Multilevel linear models

19

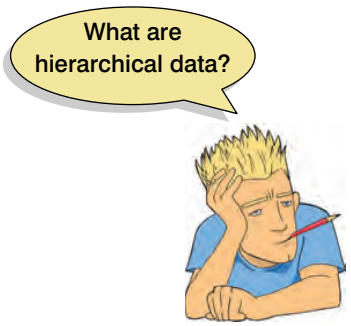
FIGURE 19.1
Having a therapy session in 2007

19.1. What will this chapter tell me? ①

Over the last couple of chapters we saw that I had gone from a child having dreams and aspirations of being a rock star, to becoming a living (barely) statistical test. A more dramatic demonstration of my complete failure to achieve my life's ambitions I can scarcely imagine. Having devoted far too much of my life to statistics it was time to unlock the latent rock star once more. The second edition of the book had left me in desperate need for some therapy and, therefore, at the age of 29 I decided to start playing the drums (there's a joke in there somewhere about it being the perfect instrument for a failed musician, but really they're much harder to play than people think). A couple of years later I had a call from an old friend of mine, Doug, who used to be in a band that my old band Scansion used to play with a lot: 'Remember the last time I saw you we talked about you coming and having a jam with us?' I had absolutely no recollection whatsoever of him saying this so I responded

'Yes'. 'Well, how about it then?' he said. 'OK,' I said, 'you arrange it and I'll bring my guitar.' 'No, you whelk,' he said, 'we want you to drum and maybe you could learn some of the songs on the CD I gave you last year?' I'd played his band's CD and I liked it, but there was no way on this earth that I could play the drums as well as their drummer. 'Sure, no problem,' I lied. I spent the next two weeks playing along to this CD as if my life depended on it and when the rehearsal came, much as I'd love to report that I drummed like a lord, I didn't. I did, however, nearly have a heart attack and herniate everything in my body that it's possible to herniate (really, the music is pretty fast!). Still, we had another rehearsal, and then another and, well, three years down the line we're still having them. The only difference is that now I can play the songs at a speed that makes their old recordings seem as though a sedated snail was on the drums (www.myspace.com/fracturepattern). The point is that it's never too late to learn something new. This is just as well because, as a man who clearly doesn't learn from his mistakes, I agreed to write a chapter on multilevel linear models, a subject about which I know absolutely nothing. I'm writing it last, when I feel mentally exhausted and stressed. Hopefully at some point between now and the end of writing the chapter I will learn something. With a bit of luck you will too.

19.2. Hierarchical data ②



In all of the analyses in this book so far we have treated data as though they are organized at a single level. However, in the real world, data are often hierarchical. This just means that some variables are clustered or *nested* within other variables. For example, when I'm not writing statistics books I spend most of my time researching how anxiety develops in children below the age of 10. This typically involves my running experiments in schools. When I run research in a school, I test children who have been assigned to different classes, and who are taught by different teachers. The classroom that a child is in could conceivably affect my results. Let's imagine I test in two different classrooms. The first class is taught by Mr Nervous. Mr Nervous is very anxious and often when he supervises children he tells them to be careful, or that things that they do are dangerous, or that they might hurt themselves. The second class is taught by Little Miss Daredevil.[1] She is very carefree and she believes that children in her class should have the freedom to explore new experiences. Therefore, she is always telling them not to be scared of things and to explore new situations. One day I go into the school to test the children. I take in a big animal carrier, which I tell them has an animal inside. I measure whether they will put their hand in the carrier to stroke the animal. Children taught by Mr Nervous have grown up in an environment where their teacher reinforces caution, whereas children taught by Miss Daredevil have been encouraged to embrace new experiences. Therefore, we might expect Mr Nervous's children to be more reluctant to put their hand in the box because of the classroom experiences that they have had. The classroom is, therefore, known as a contextual variable. In reality, as an experimenter I would be interested in a much more complicated situation. For example, I might tell some of the children that the animal is a bloodthirsty beast, whereas I tell others that the animal is friendly. Now obviously I'm expecting the information I give the children to affect their enthusiasm for stroking the animal. However, it's also possible that their classroom has an effect. Therefore, my manipulation of the information that I give the children also has to be placed within the context of the classroom to which the

[1] Those of you who don't spot the Mr Men references here, check out http://www.mrmen.com. Mr Nervous used to be called Mr Jelly and was a pink jelly-shaped blob, which in my humble opinion was better than his current incarnation.

child belongs. My threat information is likely to have more impact on Mr Nervous's children than it will on Miss Daredevil's children. One consequence of this is that children within Mr Nervous's class will be more similar to each other than they are to children in Miss Daredevil's class and vice versa.

Figure 19.2 illustrates this scenario more generally. In a big data set, we might have collected data from lots of children. This is the bottom of the hierarchy and is known as a *level 1* variable. So, children (or cases) are our level 1 variable. However, these children are organized by classroom (children are said to be *nested* within classes). The class to which a child belongs is a level up from the participant in the hierarchy and is said to be a *level 2* variable.

The situation that I have just described is the simplest hierarchy that you can have because there are just two levels. However, you can have other layers to your hierarchy. The easiest way to explain this is to stick with our example of my testing children in different classes and then to point out the obvious fact that classrooms are themselves nested within schools. Therefore, if I ran a study incorporating lots of different schools, as well as different classrooms within those schools, then I would have to add another level to the hierarchy. We can apply the same logic as before, in that children in particular schools will be more similar to each other than to children in different schools. This is because schools tend to reflect their social demographic (which can differ from school to school) and they may differ in their policies also. Figure 19.3 shows this scenario. There are now three levels in the hierarchy: the child (level 1), the class to which the child belongs (level 2) and the school within which that class exists (level 3). In this situation we have two contextual variables: school and classroom.

Hierarchical data structures need not apply only to between-participant situations. We can also think of data as being nested within people. In this situation the case, or person, is not at the bottom of the hierarchy (level 1), but is further up. A good example is memory. Imagine that after giving children threat information about my caged animal I asked them a week later to recall everything they could about the animal. For each child there are many facts that they could recall. Let's say that I originally gave them 15 pieces of information; some children might recall all 15 pieces of information, but others might remember only 2 or 3 bits of information. The bits of information, or memories, are nested within the person and their recall depends on the person. The probability of a given memory being recalled depends on what other memories are available, and the recall of one memory may have knock-on effects for what other memories are recalled. Therefore, memories are not independent units. As such, the person acts as a context within which memories are recalled (Wright, 1998).

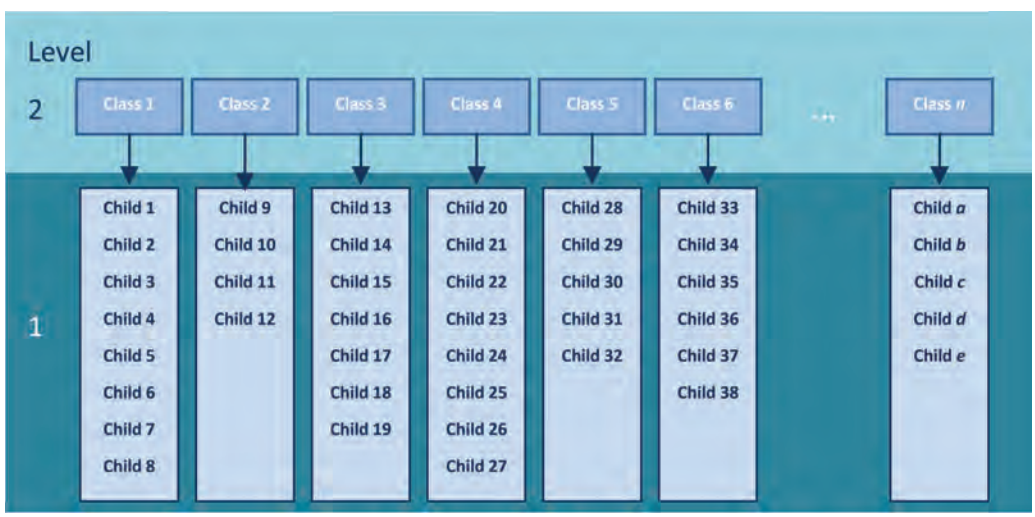


FIGURE 19.2 An example of a two-level hierarchical data structure. Children (level 1) are organized within classrooms (level 2)

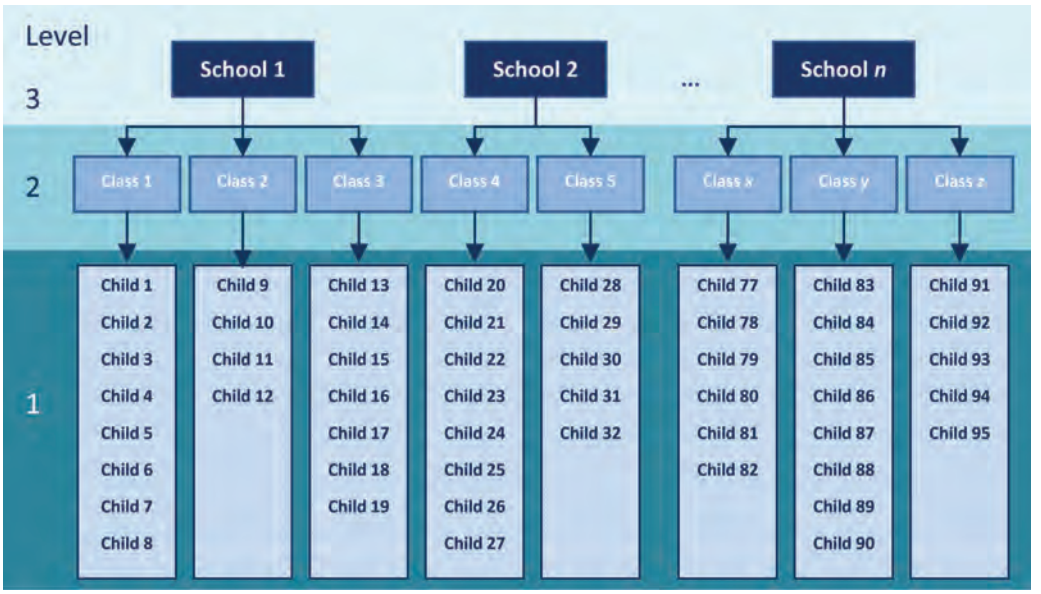


FIGURE 19.3
An example of a three-level hierarchical data structure

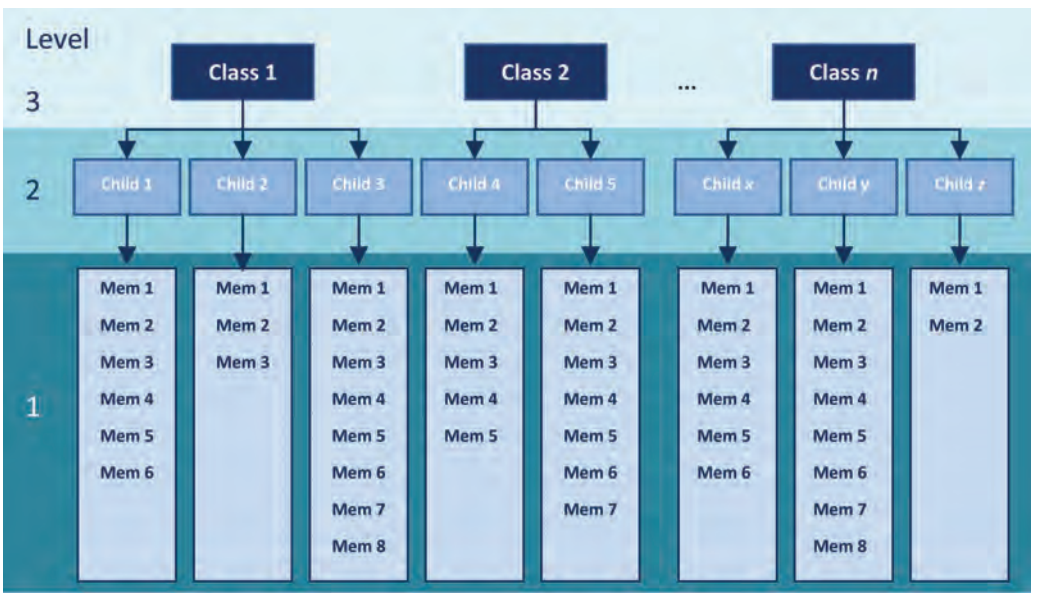


FIGURE 19.4
An example of a three-level hierarchical data structure, where the level 1 variable is a repeated measure (memories recalled)

Figure 19.4 shows the structure of the situation that I have just described. The child is our level 2 variable, and within each child there are several memories (our level 1 variable). Of course we can also have levels of the hierarchy above the child. So, we could still, for example, factor in the context of the class from which they came (as I have done in Figure 19.4) as a level 3 variable. Indeed, we could even include the school again as a level 4 variable!

19.2.1. The intraclass correlation ②

You might well wonder why it matters that data are hierarchical (or not). The main problem is that the contextual variables in the hierarchy introduce dependency in the data. In plain English this means that residuals will be correlated. I have alluded to this fact already

when I noted that children in Mr Nervous's class would be more similar to each other than to children in Miss Daredevil's class. In some sense, having the same teacher makes children more similar to each other. This similarity is a problem because in nearly every test we have covered in this book we assume that cases are independent. In other words, there is absolutely no correlation between residual scores of one child and another. However, when people are sampled from similar contexts, this independence is unlikely to be true. For example, Charlotte and Emily's responses to the animal in the carrier have both been influenced by Mr Nervous's cautious manner, so their behaviour will be similar. Likewise, Kiki and Jip's responses to the animal in the box have both been influenced by Miss Daredevil's carefree manner, so their behaviour will be similar too. We have seen before that in ANOVA, for example, a lack of independence between cases is a huge problem that really affects the resulting test statistic – and not in a good way! (See section 10.2.10.)

By thinking about contextual variables and factoring them into the analysis we can overcome this problem of non-independent observations. One way that we can do this is to use the intraclass correlation (ICC). We came across this measure in section 17.9.3 as a measure of inter-rater reliability, but it can also be used as a measure of dependency between scores. We'll skip the formalities of calculating the ICC (but see Oliver Twisted if you're keen to know), and we'll just give a conceptual grasp of what it represents. In our two-level example of children within classes, the ICC represents the proportion of the total variability in the outcome that is attributable to the classes. It follows that if a class has had a big effect on the children within it then the variability within the class will be small (the children will behave similarly). As such, variability in the outcome within classes is minimized, and variability in the outcome between classes is maximized; therefore, the ICC is large. Conversely, if the class has little effect on the children then the outcome will vary a lot within classes, which will make differences between classes relatively small. Therefore, the ICC is small too. Thus, the ICC tells us that variability within levels of a contextual variable (in this case the class to which a child belongs) is small, but between levels of a contextual variable (comparing classes) is large. As such the ICC is a good gauge of whether a contextual variable has an effect on the outcome.

OLIVER TWISTED

Please, Sir, can I have some more … ICC?

'I have a dependency on gruel,' whines Oliver. 'Maybe I could measure this dependency if I knew more about the ICC.' We'll you're so high on gruel Oliver that you have rather missed the point. Still, I did write an article on the ICC once upon a time (Field, 2005a) and it's reproduced in the additional web material for your delight and amusement.

19.2.2. Benefits of multilevel models ②

Multilevel linear models have numerous uses. To convince you that trawling through this chapter is going to reward you with statistical possibilities beyond your wildest dreams, here are just a few (slightly overstated) benefits of multilevel models:

- **Cast aside the assumption of homogeneity of regression slopes**: We saw in Chapter 11 that when we use analysis of covariance we have to assume that the relationship between our covariate and our outcome is the same across the different groups that make up our predictor variable. However, this doesn't always happen. Luckily, in multilevel models we can explicitly model this variability in regression slopes, thus overcoming this inconvenient problem.

- **Say 'bye bye' to the assumption of independence**: We saw in Chapter 10 that when we use independent ANOVA we have to assume that the different cases of data are independent. If this is not true, little lizards climb out of your mattress while you're asleep and eat you. Again, multilevel models are specifically designed to allow you to model these relationships between cases. Also, in Chapter 7 we saw that multiple regression relies on having independent observations. However, there are situations in which you might want to measure someone on more than one occasion (i.e. over time). Ordinary regression turns itself into cheese and hides in the fridge at the prospect of cases of data that are related. Multilevel models eat these data for breakfast, with a piece of regression-flavoured cheese.
- **Laugh in the face of missing data**: I've spent a lot of this book extolling the virtues of balanced designs and not having missing data. Regression, ANOVA, ANCOVA and most of the other tests we have covered do strange things when data are missing or the design is not balanced. This can be a real pain. Multilevel models open the door to missing data, invite them to sit by the fire and make them a cup of tea. Multilevel models expect missing data, they love them in fact. So, if you have some kind of ANOVA or regression (of any variety) for which you have missing data, fear not, just do a multilevel model.

I think you'll agree that multilevel models are pretty funky. 'Is there anything they can't do?' I hear you cry. Well, no, not really.

19.3. Theory of multilevel linear models ③

The underlying theory of multilevel models is very complicated indeed – far too complicated for my little peanut of a brain to comprehend. Fortunately, the advent of computers and software like SPSS makes it possible for feeble-minded individuals such as myself to take advantage of this wonderful tool without actually needing to know the maths. Better still, this means I can get away with not explaining the maths (and really, I'm not kidding, I don't understand any of it). What I will do though is try to give you a flavour of what multilevel models are and what they do by describing the key concepts within the framework of linear models that has permeated this whole book.

19.3.1. An example ②

Throughout the first part of the chapter we will use an example to illustrate some of the concepts in multilevel models. Cosmetic surgery is on the increase at the moment. In the USA, there was a 1600% increase in cosmetic surgical and non-surgical treatments between 1992 and 2002, and in 2004, 65,000 people in the UK underwent privately and publicly funded operations (Kellett, Clarke, & McGill, 2008). With the increasing popularity of this surgery, many people are starting to question the motives of those who want to go under the knife. There are two main reasons to have cosmetic surgery: (1) to help a physical problem such as having breast reduction surgery to relieve back ache; and (2) to change your external appearance, for example by having a face lift. Related to this second point, there is even some case for arguing that cosmetic surgery could be performed as a psychological intervention: to improve self-esteem (Cook, Rosser, & Salmon, 2006; Kellett et al., 2008). The main example for this chapter looks at the effects of cosmetic surgery on quality of life. The variables in the data file are:

- **Post_QoL**: This is a measure of quality of life after the cosmetic surgery. This is our outcome variable.
- **Base_QoL**: We need to adjust our outcome for quality of life before the surgery.
- **Surgery**: This variable is a dummy variable that specifies whether the person has undergone cosmetic surgery (1) or whether they are on the waiting list (0), which acts as our control group.
- **Clinic**: This variable specifies which of 10 clinics the person attended to have their surgery.
- **Age**: This variable tells us the person's age in years.
- **BDI**: It is becoming increasingly apparent that people volunteering for cosmetic surgery (especially when the surgery is purely for vanity) might have very different personality profiles than the general public (Cook, Rossera, Toone, James, & Salmon, 2006). In particular, these people might have low self-esteem or be depressed. When looking at quality of life it is important to assess natural levels of depression and this variable used the Beck Depression Inventory (BDI) to do just that.
- **Reason**: This dummy variable specifies whether the person had/is waiting to have surgery purely to change their appearance (0), or because of a physical reason (1).
- **Gender**: This variable simply specifies whether the person was a man (1) or a woman (0).

When conducting hierarchical models we generally work up from a very simple model to more complicated models and we will take that approach in this chapter. In doing so I hope to illustrate multilevel modelling by attaching it to frameworks that you already understand, such as ANOVA and ANCOVA.

Figure 19.5 shows the hierarchical structure of the data. Essentially, people being treated in the same surgeries are not independent of each other because they will have had surgery from the same surgeon. Surgeons will vary in how good they are, and quality of life will to some extent depend on how well the surgery went (if they did a nice neat job then quality of life should be higher than if they left you with unpleasant scars). Therefore, people within clinics will be more similar to each other than people in different clinics. As such, the person undergoing surgery is the level 1 variable, but there is a level 2 variable, a variable higher in the hierarchy, which is the clinic attended.

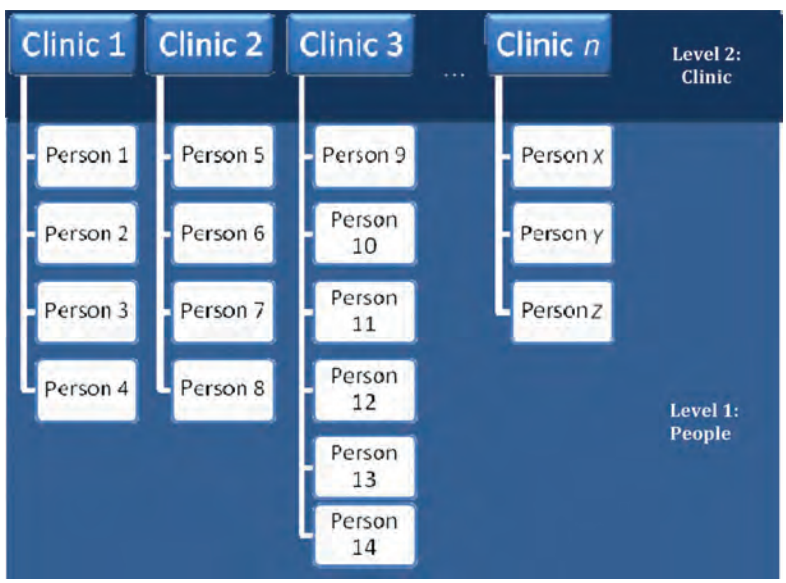


FIGURE 19.5 Diagram to show the hierarchical structure of the cosmetic surgery data set. People are clustered within clinics. Note that for each person there would be a series of variables measured: surgery, BDI, age, gender, reason and pre-surgery quality of life

19.3.2. Fixed and random coefficients ③

Throughout this book we have discussed effects and variables and these concepts should be very familiar to you by now. However, we have viewed these effects and variables in a relatively simple way: we have not distinguished between whether something is fixed or random.

What we mean by 'fixed' and 'random' can be a bit confusing because the terms are used in a variety of contexts. You hear people talk about **fixed effects** and **random effects**. An effect in an experiment is said to be a fixed effect if all possible treatment conditions that a researcher is interested in are present in the experiment. An effect is said to be random if the experiment contains only a random sample of possible treatment conditions. This distinction is important because fixed effects can be generalized only to the situations in your experiment, whereas random effects can be generalized beyond the treatment conditions in the experiment (provided that the treatment conditions are representative). For example, in our Viagra example from Chapter 10, the effect is fixed if we say that we are interested only in the three conditions that we had (placebo, low dose and high dose) and we can generalize our findings only to the situation of a placebo, low dose and high dose. However, if we were to say that the three doses were only a sample of possible doses (maybe we could have tried a very high dose), then it is a random effect and we can generalize beyond just placebos, low doses and high doses. All of the effects in this book so far we have treated as fixed effects. The vast majority of academic research that you read will treat variables as fixed effects.

People also talk about **fixed variables** and **random variables**. A fixed variable is one that is not supposed to change over time (e.g. for most people their gender is a fixed variable – it never changes), whereas a random one varies over time (e.g. your weight is likely to fluctuate over time).

In the context of multilevel models we need to make a distinction between **fixed coefficients** and **random coefficients**. In the regressions, ANOVAs and ANCOVAs throughout this book we have assumed that the regression parameters are fixed. We have seen numerous times that a linear model is characterized by two things: the intercept, b_0, and the slope, b_1:

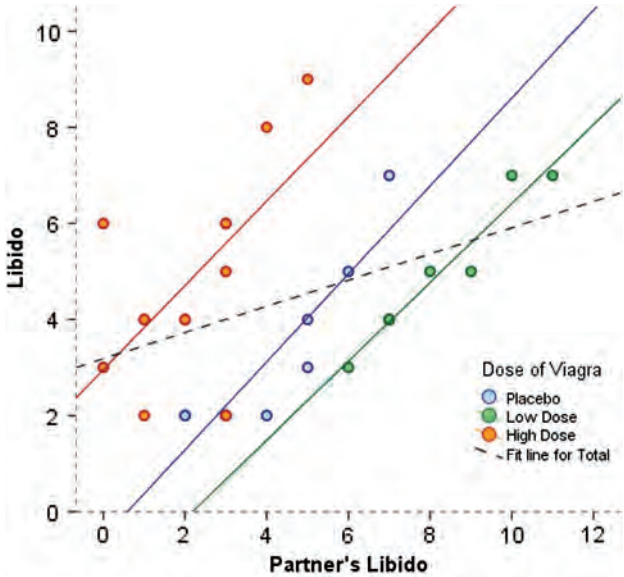
$$Y_i = b_0 + b_1 X_{1i} + \varepsilon_i$$

Note that the outcome (Y), the predictor (X) and the error (ε) all vary as a function of i, which normally represents a particular case of data. In other words, it represents the level 1 variable. If, for example, we wanted to predict Sam's score, we could replace the is with her name:

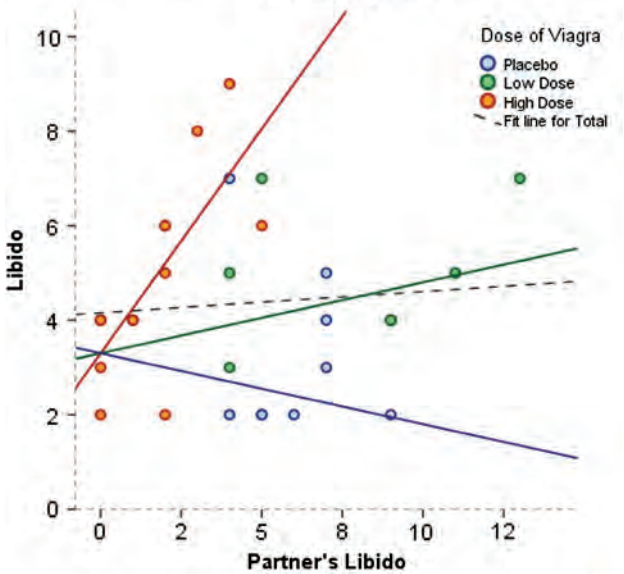
$$Y_{\text{Sam}} = b_0 + b_1 X_{1\text{Sam}} + \varepsilon_{\text{Sam}}$$

This is just some basic revision. Now, when we do a regression like this we assume that the bs are fixed and we estimate them from the data. In other words, we're assuming that the model holds true across the entire sample and that for every case of data in the sample we can predict a score using the same values of the gradient and intercept. However, we can also conceptualize these parameters as being random.[2] If we say that a parameter is random then we assume not that it is a fixed value, but that its value can vary. Up until now we have thought of regression models as having fixed intercepts and fixed slopes, but this opens up three new possibilities for us that are shown in Figure 19.6. This figure uses the data from our ANCOVA example in Chapter 11 and shows the relationship between a person's libido and that of their partner overall (the dashed line) and separately for the three groups in the study (a placebo group, a group that had a low dose of Viagra and a group that had a high dose).

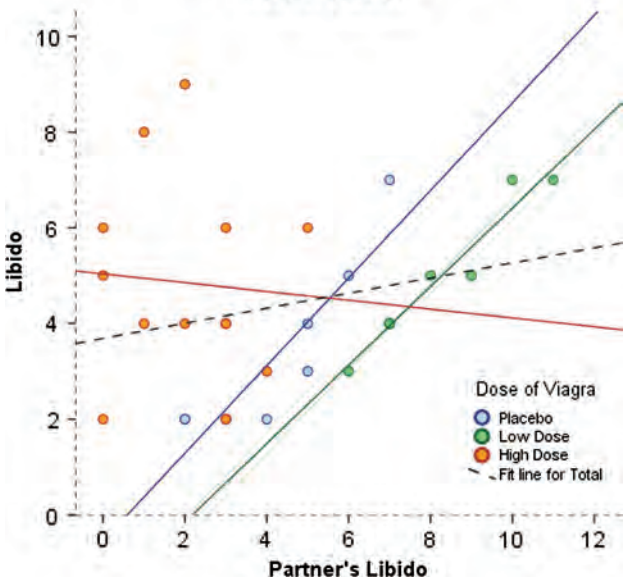
[2] In a sense random isn't an intuitive term for us non-statisticians because it implies that values are plucked out of thin air (randomly selected). However, this is not the case, they are carefully estimated just as fixed parameters are.



Random Intercept, Fixed Slope



Fixed Intercept, Random Slope



Random Intercept and Random Slope

FIGURE 19.6
Data sets showing an overall model (dashed line) and the models for separate contexts within the data (i.e. groups of cases)

19.3.2.1. The random intercept model ③

The simplest way to introduce random parameters into the model is to assume that the intercepts vary across contexts (or groups) – because the intercepts vary, we call them random intercepts. For our libido data this is like assuming that the relationship between libido and partner's libido is the same in the placebo, low- and high-dose groups (i.e. the slope is the same), but that the models for each group are in different locations (i.e. the intercepts are different). This is shown in the diagram in which the models within the different contexts (colours) have the same shape (slope) but are located in different geometric space (they have different intercepts – top panel of Figure 19.6).

19.3.2.2. Random slope model ③

We can also assume that the slopes vary across contexts – i.e. we assume random slopes. For our libido data this is like assuming that the relationship between libido and partner's libido is different in the placebo, low- and high-dose groups (i.e. the slopes are different), but that the models for each group are fixed at the same geometric location (i.e. the intercepts are the same). This is what happens when we violate the assumption of homogeneity of regression slopes in ANCOVA. Homogeneity of regression slopes is the assumption that regression slopes are the same across contexts. If this assumption is not tenable than we can use a multilevel model to explicitly estimate that variability in slopes. This is shown in the diagram in which the models within the different contexts (colours) converge on a single intercept but have different slopes (middle panel of Figure 19.6).

19.3.2.3. The random intercept and slope model ③

The most realistic situation is to assume that both intercepts and slopes vary around the overall model. This is shown in the diagram in which the models within the different contexts (colours) have different slopes but are also located in different geometric space and so have different intercepts (bottom panel of Figure 19.6).

19.4. The multilevel model ④

We have seen conceptually what a random intercept, random slope and random intercept and slope model looks like. Now let's look at how we actually represent the models. To keep things concrete, let's use our example. For the sake of simplicity, let's imagine first that we wanted to predict someone's quality of life (QoL) after cosmetic surgery. We can represent this as a linear model as follows:

$$\text{QoL After Surgery}_i = b_0 + b_1\text{Surgery}_i + \varepsilon_i \tag{19.1}$$

We have seen equations like this many times and it represents a linear model: regression, a *t*-test (in this case) and ANOVA. In this example, we had a contextual variable, which was the clinic in which the cosmetic surgery was conducted. We might expect the effect of surgery on quality of life to vary as a function of which clinic the surgery was conducted at because surgeons will differ in their skill. This variable is a level 2 variable. As such we could allow the model that represents the effect of surgery on quality of life to vary across the different contexts (clinics). We can do this by allowing the intercepts to vary across clinics, or by allowing the slopes to vary across clinics or by allowing both to vary across clinics.

To begin with, let's say we want to include a random intercept for quality of life. All we do is add a component to the intercept that measures the variability in intercepts, u_{0j}. Therefore, the intercept changes from b_0 to become $b_0 + u_{0j}$. This term estimates the intercept of the overall model fitted to the data, b_0, and the variability of intercepts around that overall model, u_{0j}. The overall model becomes.[3]

$$Y_{ij} = (b_0 + u_{0j}) + b_1 X_{ij} + \varepsilon_{ij} \tag{19.2}$$

The *j*s in the equation reflect levels of the variable over which the intercept varies (in this case the clinic) – the level 2 variable. Another way that we could write this is to take out the error terms so that it looks like an ordinary regression equation except that the intercept has changed from a fixed, b_0, to a random one, b_{0j}, which is defined in a separate equation:

$$\begin{aligned} Y_{ij} &= b_{0j} + b_1 X_{ij} + \varepsilon_{ij} \\ b_{0j} &= b_0 + u_{0j} \end{aligned} \tag{19.3}$$

Therefore, if we want to know the estimated intercept for Clinic 7, we simply replace the *j* with 'clinic 7' in the second equation:

$$b_{0\text{Clinic7}} = b_0 + u_{0\text{Clinic7}}$$

If we want to include random slopes for the effect of surgery on quality of life, then all we do is add a component to the slope of the overall model that measures the variability in slopes, u_{1j}. Therefore, the gradient changes from b_1 to become $(b_1 + u_{1j})$. This term estimates the slope of the overall model fitted to the data, b_1, and the variability of slopes in different contexts around that overall model, u_{1j}. The overall model becomes (compare to the random intercept model above):

$$Y_{ij} = b_0 + (b_1 + u_{1j}) X_{ij} + \varepsilon_{ij} \tag{19.4}$$

Again we can take the error terms out into a separate equation to make the link to a familiar linear model even clearer. It now looks like an ordinary regression equation except that the slope has changed from a fixed, b_1, to a random one, b_{1j}, which is defined in a separate equation:

$$\begin{aligned} Y_{ij} &= b_{0i} + b_{1j} X_{ij} + \varepsilon_{ij} \\ b_{1j} &= b_1 + u_{1j} \end{aligned} \tag{19.5}$$

If we want to model a situation with random slopes *and* intercepts, then we combine the two models above. We still estimate the intercept and slope of the overall model (b_0 and b_1) but we also include the two terms that estimate the variability in intercepts, u_{0j}, and slopes, u_{1j}. The overall model becomes (compare to the two models above):

$$Y_{ij} = (b_0 + u_{0j}) + (b_1 + u_{1j}) X_{ij} + \varepsilon_{ij} \tag{19.6}$$

We can link this more directly to a simple linear model if we take some of these extra terms out into separate equations. We could write this model as a basic linear model, except

[3] Some people use gamma (γ), not *b*, to represent the parameters, but I prefer *b* because it makes the link to the other linear models that we have used in this book clearer.

we've replaced our fixed intercept and slope (b_0 and b_1) with their random counterparts (b_{0j} and b_{1j}):

$$\begin{aligned} Y_{ij} &= b_{0j} + b_{1j}X_{ij} + \varepsilon_{ij} \\ b_{0j} &= b_0 + u_{0j} \\ b_{1j} &= b_1 + u_{1j} \end{aligned} \tag{19.7}$$

The take-home point is that we're not doing anything terribly different from the rest of the book: it's basically just a posh regression.

Now imagine we wanted to add in another predictor, for example quality of life before surgery. Knowing what we do about multiple regression we shouldn't be invading the personal space of the idea that we can simply add this variable in with an associated beta:

$$\text{QoL After Surgery}_i = b_0 + b_1\text{Surgery}_i + b_2\text{QoL Before Surgery}_i + \varepsilon_i \tag{19.8}$$

This is all just revision of ideas from earlier in the book. Remember also that the *i* represents the level 1 variable, in this case the people we tested. Therefore, we can predict a given person's quality of life after surgery by replacing the *i* with their name:

$$\text{QoL After}_{\text{Sam}} = b_0 + b_1\text{Surgery}_{\text{Sam}} + b_2\text{QoL Before}_{\text{Sam}} + \varepsilon_{\text{Sam}}$$

Now, if we want to allow the intercept of the effect of surgery on quality of life after surgery to vary across contexts then we simply replace b_0 with b_{0j}. If we want to allow the slope of the effect of surgery on quality of life after surgery to vary across contexts then we replace b_1 with b_{1j}. So, even with a random intercept and slope, our model stays much the same:

$$\begin{aligned} \text{QoL After}_{ij} &= b_{0j} + b_{1j}\text{Surgery}_{ij} + b_2\text{QoL Before}_{ij} + \varepsilon_{ij} \\ b_{0j} &= b_0 + u_{0j} \\ b_{1j} &= b_1 + u_{1j} \end{aligned} \tag{19.9}$$

Remember that the *j* in the equation relates to the level 2 contextual variable (clinic in this case). So, if we wanted to predict someone's score we wouldn't just do it from their name, but also from the clinic they attended. Imagine our guinea pig Sam had her surgery done at Clinic 7; then we could replace the *i*s and *j*s as follows:

$$\begin{aligned} \text{QoL After Surgery}_{\text{Sam, Clinic7}} = {} & b_{0\text{Clinic7}} + b_{1\text{Clinic7}}\text{Surgery}_{\text{Sam, Clinic7}} \\ & + b_2\text{QoL Before Surgery}_{\text{Sam, Clinic7}} + \varepsilon_{\text{Sam, Clinic7}} \end{aligned}$$

I want to sum up by just reiterating that all we're really doing in a multilevel model is a fancy regression in which we allow either the intercepts or slopes, or both, to vary across different contexts. All that really changes is that for every parameter that we allow to be random, we get an estimate of the variability of that parameter as well as the parameter itself. So, there isn't anything terribly complicated; we can add new predictors to the model and for each one decide whether its regression parameter is fixed or random.

19.4.1. Assessing the fit and comparing multilevel models ④

As in logistic regression (Chapter 8) the overall fit of a multilevel model is tested using a chi-square likelihood ratio test (see section 18.3.3) and just as in logistic regression, SPSS reports the –2 log-likelihood (see section 8.3.1). Essentially, the smaller the value of the log-likelihood, the better. SPSS also produces four adjusted versions of the log-likelihood value. All of these can be interpreted in the same way as the log-likelihood, but they have been corrected for various things:

- *Akaike's information criterion* (AIC): This is basically a goodness-of-fit measure that is corrected for model complexity. That just means that it takes into account how many parameters have been estimated.
- *Hurvich and Tsai's criterion* (AICC): This is the same as AIC but is designed for small samples.
- *Bozdogan's criterion* (CAIC): Again this can be interpreted in the same way as the AIC, but this version corrects not just for model complexity but for sample size too.
- *Schwarz's Bayesian criterion* (BIC): This statistic is again comparable to the AIC, although it is slightly more conservative (it corrects more harshly for the number of parameters being estimated). It should be used when sample sizes are large and the number of parameters is small.

All of these measures are similar but the AIC and BIC are the most commonly used. None of them are intrinsically interpretable (it's not meaningful to talk about their values being large or small per se); however, they are all useful as a way of comparing models. The value of AIC, AICC, CAIC and BIC can all be compared to their equivalent values in other models. In all cases smaller values mean better-fitting models.

Many writers recommend building up multilevel models starting with a 'basic' model in which all parameters are fixed and then adding in random coefficients as appropriate and exploring confounding variables (Raudenbush & Bryk, 2002; Twisk, 2006). One advantage of doing this is that you can compare the fit of the model as you make parameters random, or as you add in variables. To compare models we simply subtract the log-likelihood of the new model from the value for the old:

$$\begin{aligned} \chi^2_{\text{Change}} &= (-2\text{Log} - \text{Likelihood}_{\text{Old}}) | (-2\text{Log} - \text{Likelihood}_{\text{New}}) \\ df_{\text{Change}} &= \text{Number of Parameters}_{\text{Old}} - \text{Number of Parameters}_{\text{New}} \end{aligned} \tag{19.10}$$

This equation is the same as equations (18.5) and (8.6), but written in a way that uses the names of the actual values that SPSS produces. There are two caveats to this equation: (1) it works only if full maximum-likelihood estimation is used (and not restricted maximal likelihood, see SPSS Tip 19.1); and (2) the new model contains all of the effects of the older model.

19.4.2. Types of covariance structures ④

If you have any random effects or repeated measures in your multilevel model then you have to decide upon the *covariance structure* of your data. If you have random effects and

repeated measures then you can specify different covariance structures for each. The covariance structure simply specifies the form of the variance–covariance matrix (a matrix in which the diagonal elements are variances and the off-diagonal elements are covariances). There are various forms that this matrix could take and we have to tell SPSS what form we think it *does* take. Of course we might not know what form it takes (most of the time we'll be taking an educated guess), so it is sometimes useful to run the model with different covariance structures defined and use the goodness-of-fit indices (the AIC, AICC, CAIC and BIC) to see whether changing the covariance structure improves the fit of the model (remember that a smaller value of these statistics means a better-fitting model).

The covariance structure is important because SPSS uses it as a starting point to estimate the model parameters. As such, you will get different results depending on which covariance structure you choose. If you specify a covariance structure that is too simple then you are more likely to make a Type I error (finding a parameter is significant when in reality it is not), but if you specify one that is too complex then you run the risk of a Type II error (finding parameters to be non-significant when in reality they are). SPSS has 17 different covariance structures that you can use. We will look at four of the commonest covariance structures to give you a feel for what they are and when they should be used. In each case I use a representation of the variance–covariance matrix to illustrate. With all of these matrices you could imagine that the rows and columns represents four different clinics in our cosmetic surgery data:

$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	**Variance components**: This covariance structure is very simple and assumes that all random effects are independent (this is why all of the covariances in the matrix are 0). Variances of random effects are assumed to be the same (hence why they are 1 in the matrix) and sum to the variance of the outcome variable. In SPSS this is the default covariance structure for random effects and is sometimes called the independence model.
$\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & 0 \\ 0 & 0 & 0 & \sigma_1^2 \end{pmatrix}$	**Diagonal**: This variance structure is like variance components except that variances are assumed to be heterogeneous (this is why the diagonal of the matrix is made up of different variance terms). This structure again assumes that variances are independent and, therefore, that all of the covariances are 0. In SPSS this is the default covariance structure for repeated measures.
$\begin{pmatrix} 1 & \rho & \rho^2 & \rho^2 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^2 & \rho^2 & \rho & 1 \end{pmatrix}$	**AR(1)**: This stands for first-order autoregressive structure. In layman's terms this means that the relationship between variances changes in a systematic way. If you imagine the rows and columns of the matrix to be points in time, then it assumes that the correlations between repeated measurements is highest at adjacent time points. So, in the first column, the correlation between time points 1 and 2 is ρ; let's assume that this value is .3. As we move to time point 3, the correlation between time point 1 and 3 is ρ^2, or .09. In other words, it has decreased: scores at time point 1 are more related to scores at time 2 than they are to scores at time 3. At time 4, the correlation goes down again to ρ^3 or .027. So, the correlations between time points next to each other are assumed to be ρ, scores two intervals apart are assumed to have correlations of ρ^2, and scores three intervals apart are assumed to have correlations of ρ^3. So the correlation between scores gets smaller over time. Variances are assumed to be homogeneous but there is a version of this covariance structure where variance can be heterogeneous. This structure is often used for repeated-measures data (especially when measurements are taken over time such as in growth models).
$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$	**Unstructured**: This covariance structure is completely general and is, therefore, the default option for random effects in SPSS. Covariances are assumed to be completely unpredictable: they do not conform to a systematic pattern.

CRAMMING SAM'S TIPS Multilevel models

- Multilevel models should be used to analyse data that have a hierarchical structure. For example, you might measure depression after psychotherapy. In your sample, patients will see different therapists within different clinics. This is a three-level hierarchy with depression scores from patients (level 1), nested within therapists (level 2) who are themselves nested within clinics (level 3).
- Hierarchical models are just like regression, except that you can allow parameters to vary (this is called a random effect). In ordinary regression, parameters generally are a fixed value estimated from the sample (a fixed effect).
- If we estimate a linear model within each context (e.g. the therapist or clinic to use the example above) rather than the sample as a whole, then we can assume that the intercepts of these models vary (a random intercepts model), or that the slopes of these models differ (a random slopes model) or that both vary.
- We can compare different models (assuming that they differ in only one additional parameter) by looking at the difference in the –2 log-likelihood. Usually we would do this when we have changed only one parameter (added one new thing to the model).
- For any model we have to assume a covariance structure. For random intercepts models the default of *variance components* is fine, but when slopes are random an *unstructured* covariance structure is often assumed. When data are measured over time an autoregressive structure (AR1) is often assumed.

19.5. Some practical issues ③

19.5.1. Assumptions ③

Multilevel linear models are an extension of regression so all of the assumptions for regression apply to multilevel models (see section 7.6.2). There is a caveat, though, which is that the assumptions of independence and independent errors can sometimes be solved by a multilevel model because the purpose of this model is to factor in the correlations between cases caused by higher-level variables. As such, if a lack of independence is being caused by a level 2 or level 3 variable then a multilevel model should make this problem go away (although not always). As such, try to check the usual assumptions in the usual way.

There are two additional assumptions in multilevel models that relate to the random coefficients. These coefficients are assumed to be normally distributed around the overall model. So, in a random intercepts model the intercepts in the different contexts are assumed to be normally distributed around the overall model. Similarly, in a random slopes model, the slopes of the models in different contexts are assumed to be normally distributed.

Also it's worth mentioning that multicollinearity can be a particular problem in multilevel models if you have interactions that cross levels in the data hierarchy (cross-level interactions). However, centring predictors can help matters enormously (Kreft & de Leeuw, 1998), and we will see how to centre predictors in section 19.5.3.

19.5.2. Sample size and power ③

As you might well imagine, the situation with power and sample size is very complex indeed. One complexity is that we are trying to make decisions about our power to detect both fixed and random effects coefficients. Kreft and de Leeuw (1998) do a tremendous job of making sense of things for us. Essentially, the take-home message is the more data, the better. As more levels are introduced into the model, more parameters need to be estimated and the larger the sample sizes need to be. Kreft and de Leeuw conclude that if you are looking for cross-level interactions then you should aim to have more than 20 contexts (groups) in the higher-level variable, and that group sizes 'should not be too small'. They conclude by saying that there are so many factors involved in multilevel analysis that it is impossible to produce any meaningful rules of thumb.

Twisk (2006) agrees that the number of contexts relative to individuals within those contexts is important. He also points out that standard sample size and power calculations can be used but then 'corrected' for the multilevel component of the analysis (by factoring, among other things, the intraclass correlation). However, there are two corrections that he discusses that yield very different sample sizes! He recommends using sample size calculations with caution.

The easiest option is to get a computer to do it for you. HLM (http://www.ssicentral.com/hlm/index.html) will do power calculations for multilevel models, and for two-level models you could try Tom Snijders' PinT program (http://stat.gamma.rug.nl/multilevel.htm).

19.5.3. Centring variables ④

Centring refers to the process of transforming a variable into deviations around a fixed point. This fixed point can be any value that you choose, but typically we use the grand mean. We have already come across a form of centring way back in Chapter 1, when we discovered how to compute *z*-scores. When we calculate a *z*-score we take each score and subtract from it the mean of all scores (this centres the values at 0), and then divide by the standard deviation (this changes the units of measurement to standard deviations). When we centre a variable around the mean we simply subtract the mean from all of the scores: this centres the variables around 0.

There are two forms of centring that are typically used in multilevel modelling: **grand mean centring** and **group mean centring**. Grand mean centring means that for a given variable we take each score and subtract from it the mean of all scores (for that variable). Group mean centring means that for a given variable we take each score and subtract from it the mean of the scores (for that variable) within a given group. In both cases it is usually only level 1 predictors that are centred (in our cosmetic surgery example this would be predictors such as age, BDI and pre-surgery quality of life). If group mean centring is used then a level 1 variable is typically centred around means of a level 2 variable (in our cosmetic surgery data this would mean that, for example, the age of a person would be centred around the mean of age for the clinic at which the person had their surgery).

Centring can be used in ordinary multiple regression too, and because this form of regression is already familiar to you I'd like to begin by looking at the effects of centring in regression. In multiple regression the intercept represents the value of the outcome when all of the predictors take a value of 0. There are some predictors for which a value of 0 makes little sense. For example, if you were using heart rate as a predictor variable then a value of 0 would be meaningless (no one will have a heart rate of 0 unless they are dead). As such, the

intercept in this case has no real-world use: why would you want to know the value of the outcome when heart rate was 0 given than no alive person would even have a heart rate that low? Centring heart rate around its mean changes the meaning of the intercept. The intercept becomes the value of the outcome when heart rate is its average value. In more general terms, if all predictors are centred around their mean then the intercept is the value of the outcome when all predictors are the value of their mean. Centring can, therefore, be a useful tool for interpretation when a value of 0 for the predictor is meaningless.

The effect of centring in multilevel models, however, is much more complicated. There are some excellent reviews that look in detail at the effects of centering on multilevel models (Kreft & de Leeuw, 1998; Kreft, de Leew, & Aiken, 1995), and here I will just give a very basic précis of what they say. Essentially if you fit a multilevel model using the raw score predictors and then fit the same model but with grand mean centred predictors then the resulting models are equivalent. By this, I mean that they will fit the data equally well, have the same predicted values, and the residuals will be the same. The parameters themselves (the *b*s) will, of course, be different but there will be a direct relationship between the parameters from the two models (i.e. they can be directly transformed into each other). Therefore, grand mean centring doesn't change the model, but it would change your interpretation of the parameters (you can't interpret them as though they are raw scores). When group mean centring is used the picture is much more complicated. In this situation the raw score model is not equivalent to the centred model in either the fixed part or the random part. One exception is when only the intercept is random (which arguably is an unusual situation), and the group means are reintroduced into the model as level 2 variables (Kreft & de Leeuw, 1998).

The decision about whether to centre or not is quite complicated and you really need to make the decision yourself in a given analysis. Centring can be a useful way to combat multicollinearity between predictor variables. It's also helpful when predictors do not have a meaningful zero point. Finally, multilevel models with centred predictors tend to be more stable, and estimates from these models can be treated as more or less independent of each other, which might be desirable. If group mean centring is used then the group means should be reintroduced as a level 2 variable unless you want to look at the effect of your 'group' or level 2 variable uncorrected for the mean effect of the centred level 1 predictor, such as when fitting a model when time is your main explanatory variable (Kreft & de Leeuw, 1998).

OLIVER TWISTED

Please, Sir, can I have some more … centring?

'Recentgin' babbles Oliver as he stumbles drunk out of Mrs Moonshine's alcohol emporium. 'I need some more recent gin.' I think you mean *centring* Oliver, not *recentgin*. If you want to know how to centre your variables using SPSS, then the additional material for this chapter on the companion website will tell you.

19.6. Multilevel modelling on SPSS ④

SPSS is not the best program in the world for multilevel modelling. Most people who do serious multilevel modelling tend to use specialist software such as MLwiN, HLM, SAS and R. There are several excellent books that compare the various packages and SPSS tends to fare pretty badly in all of them (Tabachnick & Fidell, 2001; Twisk, 2006). The main area where SPSS is behind its competitors is that it cannot do multilevel modelling when the outcome variable is categorical, yet this is bread and butter (albeit staggeringly complicated bread and butter) for the other packages mentioned. The second problem is that SPSS cannot produce

bootstrap estimates of the model parameters, and these can be a very useful way to circumvent pesky distributional assumptions (see section 5.7.4). Other packages have these facilities. SPSS also has (and it's not just me that says this) a completely indecipherable windows interface for doing multilevel models (it is much easier to do using syntax).

We saw in section 19.4.1 that it is useful to build up models starting with a 'basic' model in which all parameters are fixed and then add random coefficients as appropriate before exploring confounding variables. We will take this approach to look at an example of conducting a multilevel model on SPSS.

19.6.1. Entering the data ②

Data entry depends a bit on the type of multilevel model that you wish to run: the data layout is slightly different when the same variables are measured at several points in time. However, we will look at the case of repeated-measures data in a second example. In this first example, the situation we have is very much like multiple regression in that data from each person who had surgery are not measured over multiple time points. Figure 19.7 shows the data layout. Each row represents a case of data (in this case a person who had surgery). Their scores on the various variables are simply entered in different columns. So, for example, the first person was 31 years old, had a BDI score of 12, they were in the waiting list control group at clinic 1, were female and were waiting for surgery for a physical reason.

19.6.2. Ignoring the data structure: ANOVA ②

First of all, let's ground the example in something very familiar to us: ANOVA. Let's say for the time being that we were interested only in the effect that surgery has on post-operative quality of life. We could analyse this with a simple one-way independent ANOVA (or indeed a *t*-test), and the model is described by equation (19.1).

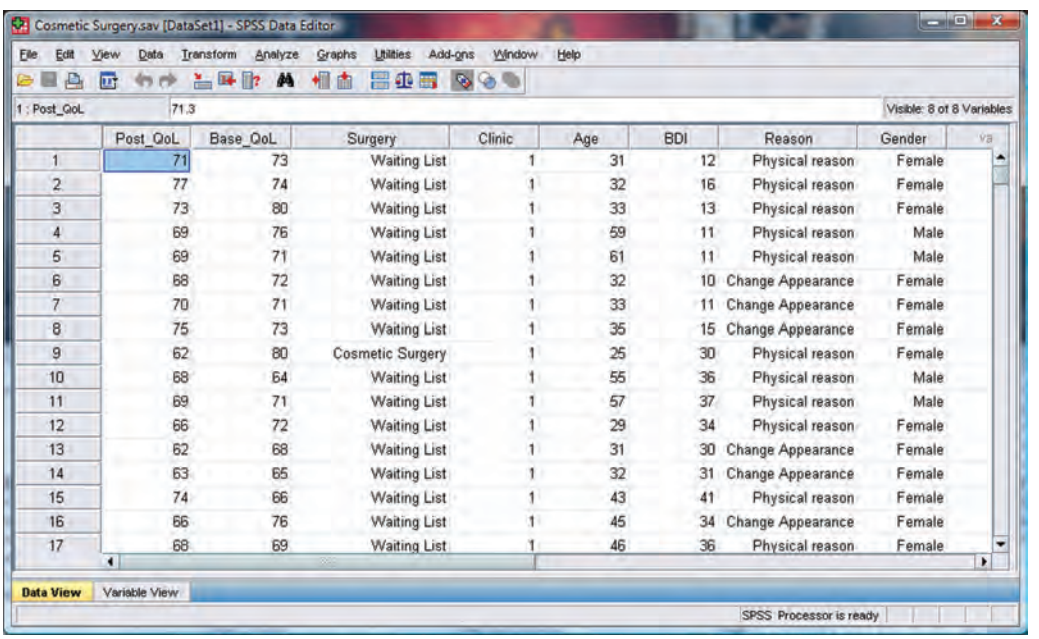


FIGURE 19.7 Data layout for multilevel modelling with no repeated measure

SELF-TEST Using what you know about ANOVA, conduct a one-way ANOVA using **Surgery** as the predictor and **Post_QoL** as the outcome.

In reality we wouldn't do an ANOVA, I'm just using it as a way of showing you that multilevel models are not big and scary, but are simply extensions of what we have done before. SPSS Output 19.1 shows the results of the ANOVA that you should get if you did the self-test. We find a non-significant effect of surgery on quality of life, $F(1, 274) = 0.33$, $p > .05$.

SPSS OUTPUT 19.1

ANOVA

Quality of Life After Cosmetic Surgery

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28.620	1	28.620	.330	.566
Within Groups	23747.883	274	86.671		
Total	23776.504	275			

To run a multilevel model we use the *Mixed Models* command. To access this command select Analyze Mixed Models Linear..., which will bring up the dialog box in Figure 19.8. This dialog box is for specifying the hierarchical nature of the data and because for the time being we are ignoring the hierarchical structure of our data, we will ignore this dialog box for now.

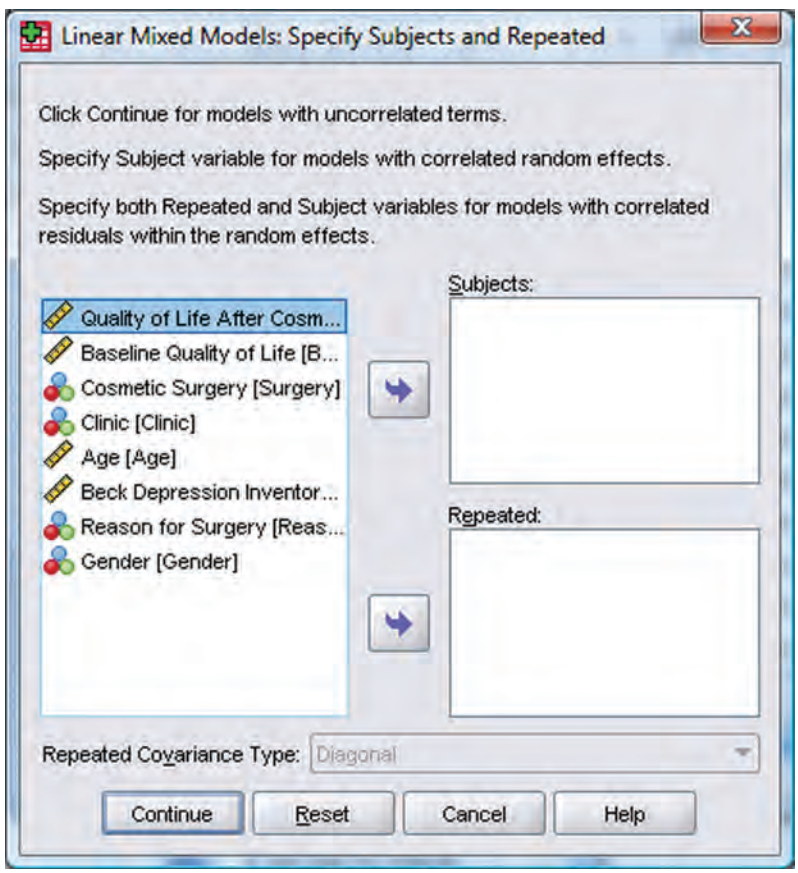


FIGURE 19.8 The initial mixed models dialog box

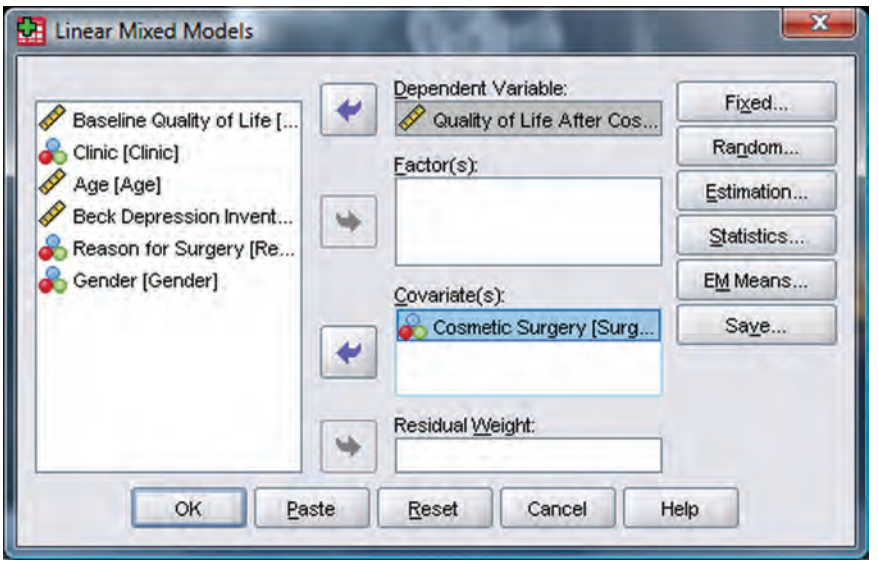


FIGURE 19.9
The main mixed models dialog box

Click on Continue to move to the main dialog box (Figure 19.9), which should look very familiar to many other dialog boxes that we have seen before. First we must specify our outcome variable, which is quality of life (QoL) after surgery, so select **Post_QoL** and drag it to the space labelled *Dependent Variable* (or click on). Next we need to specify our predictor, which is whether or not the person has had surgery. Therefore, select **Surgery** and drag it to the space labelled *Covariate(s)* (or click on).[4]

You'll notice several buttons at the side of the main dialog box. We use Fixed... to specify fixed effects in our model, and Random... to specify, yes, you've guessed it, random effects. To begin with we are going to treat our effects as fixed, so click on Fixed... to bring up the dialog box in Figure 19.10. We have only one variable specified as a predictor, and we want this to be treated as a fixed effect; therefore, we select it in this dialog box from the list labelled *Factors and Covariates* and then click on Add to transfer it to the *Model*. Click on Continue to return to the main dialog box.

In the main dialog box click on Estimation... to open the dialog box in Figure 19.11 (left panel). This dialog box allows you to change the parameters that SPSS will use when estimating the model. For example, if you don't get a solution then you could increase the number of iterations (SPSS Tip 8.1). The defaults can be left alone, but you do need to decide whether to use the maximum likelihood, or something called the restricted maximum-likelihood estimation method. There are pros and cons to both (see SPSS Tip 19.1) but because we want to compare models as we build them up, we will select Maximum Likelihood (ML). Click on Continue to return to the main dialog box.

In the main dialog box click on Statistics... to open the dialog box in Figure 19.11 (right panel). There are two useful options in this dialog box. The first is to request *Parameter estimates*. This will give us *b*-values for each effect and their significance (so, it will give us similar information to the coefficients table in multiple regression). The second useful option is *Tests for covariance parameters*, which will give us a significance test of each of the covariance estimates in the model (i.e. the values of u in equations (19.3), (19.5) and

[4] You might wonder why we don't drag it to the *Factors* box given that it is a categorical variable. I wondered that too, but when I did drag it there the resulting analysis is wrong. Given this variable is coded 0 and 1 it shouldn't make any difference whether we specify it as a covariate of a factor, but when we include the hierarchical data structure it does. I don't know why, maybe you can email SPSS and then tell me.

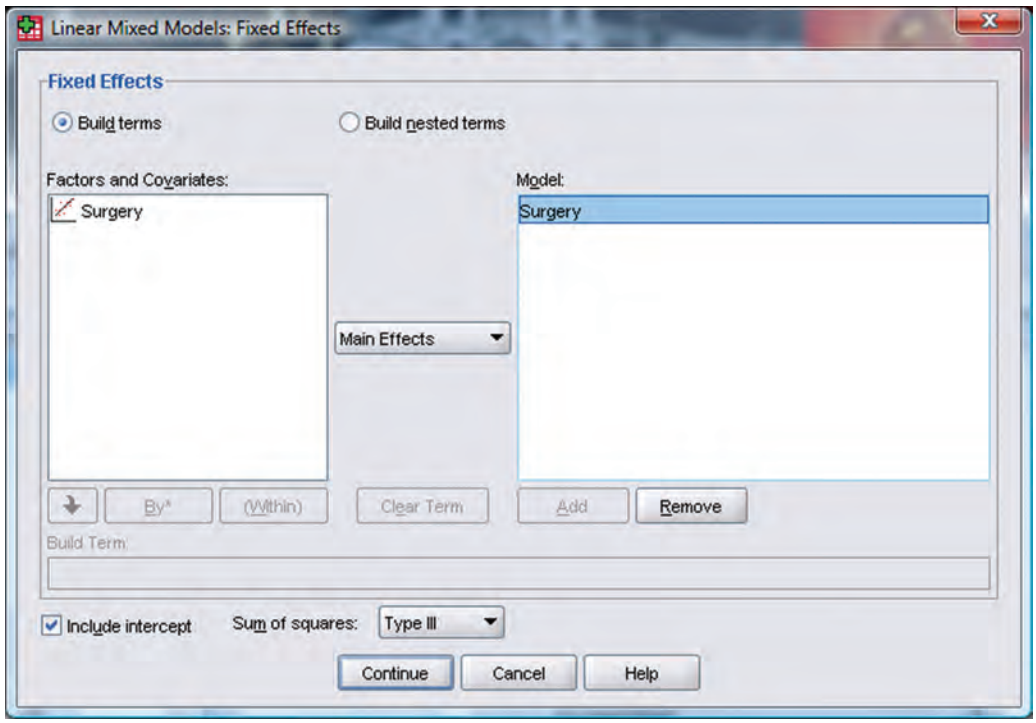
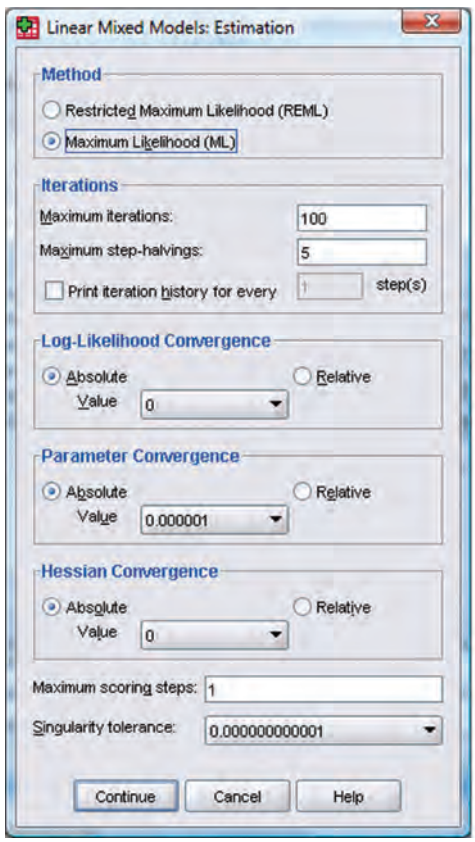


FIGURE 19.10
The dialog box for specifying fixed effects in mixed models



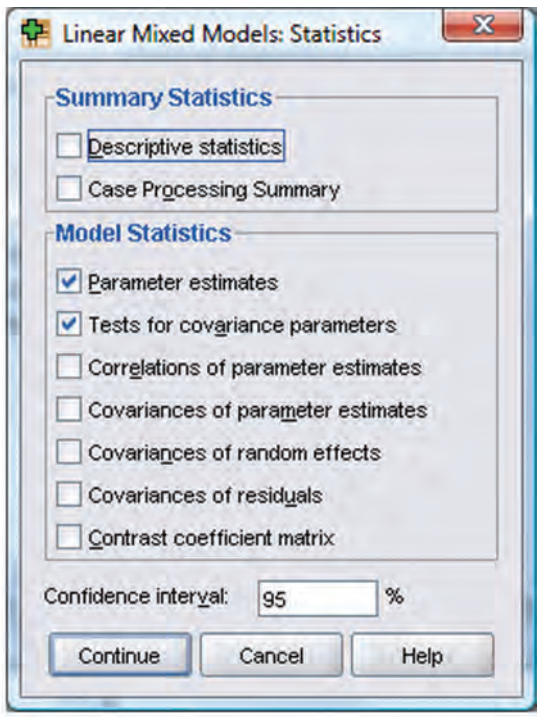


FIGURE 19.11
The estimation and statistics options for mixed models

SPSS TIP 19.1 Estimation ③

SPSS gives you the choice of two methods for estimating the parameters in the analysis: maximum likelihood (ML), which we have encountered before, and restricted maximum likelihood (REML). The conventional wisdom seems to be that ML produces more accurate estimates of fixed regression parameters, whereas REML produces more accurate estimates of random variances (Twisk, 2006). As such, the choice of estimation procedure depends on whether your hypotheses are focused on the fixed regression parameters or on estimating variances of the random effects. However, in many situations the choice of ML or REML will make only small differences to the parameter estimates. Also, if you want to compare models you must use ML.

(19.7)). These estimates tell us about the variability of intercepts or slopes across our contextual variable and so significance testing them can be useful (we can then say that there was significant, or not, variability in intercepts or slopes). Select these two options and then click on Continue to return to the main dialog box. To run the analysis, click on OK.

SPSS Output 19.2 shows the main table for the model. Compare this table with SPSS Output 19.1 and you'll see that there is basically no difference: we get a non-significant effect of surgery with an F of 0.33, and a p of .56. The point I want you to absorb here is that if we ignore the hierarchical structure of the data then what we are left with is something very familiar: an ANOVA/regression. The numbers are more or less exactly the same; all that has changed is that we have used different menus to get to the same end point.

SPSS OUTPUT 19.2

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	276	6049.727	.000
Surgery	1	276	.333	.565

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

19.6.3. Ignoring the data structure: ANCOVA ②

We have seen that there is no effect of cosmetic surgery on quality of life, but we did not take into account the quality of life before surgery. Let's, therefore, extend the example a little to look at the effect of the surgery on quality of life while taking into account the quality of life scores before surgery. Our model is now described by equation (19.8). You would normally do this analysis with an ANCOVA, through the univariate GLM menu. As in the previous section we'll run the analysis both ways, just to illustrate that we're doing the same thing when we run a hierarchical model.

SELF-TEST Using what you know about ANCOVA, conduct a one-way ANCOVA using **Surgery** as the predictor, **Post_QoL** as the outcome and **Base_QoL** as the covariate.

As before, we probably wouldn't do an ANCOVA using the mixed model command, but it's a useful illustration. SPSS Output 19.3 shows the results of the ANCOVA that you should get if you did the self-test. With baseline quality of life included we find a significant effect of surgery on quality of life, $F(1, 273) = 4.04$, $p < .05$. Baseline quality of life also predicted quality of life after surgery, $F(1, 273) = 214.89$, $p < .001$.

SPSS OUTPUT 19.3

Tests of Between-Subjects Effects

Dependent Variable:Quality of Life After Cosmetic Surgery

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10488.253[a]	2	5244.127	107.738	.000
Intercept	1713.257	1	1713.257	35.198	.000
Base_QoL	10459.633	1	10459.633	214.888	.000
Surgery	196.816	1	196.816	4.043	.045
Error	13288.250	273	48.675		
Total	1004494.530	276			
Corrected Total	23776.504	275			

a. R Squared = .441 (Adjusted R Squared = .437)

Select Analyze Mixed Models Linear... again, and just like last time ignore the first dialog box because for the time being we are ignoring the hierarchical structure of our data. We can leave the main dialog box (Figure 19.12) as it was in the last analysis except that we now need to add the baseline quality of life as another predictor. To do this, select **Base_QoL** and drag it to the space labelled *Covariate(s)* (or click on).

We need to add this new variable to our model as a fixed effect, so click on Fixed... to bring up the dialog box in Figure 19.13. Select **Base_QoL** in the list labelled *Factors and Covariates* and then click on Add to transfer it to the *Model*. Click on Continue to return to the main dialog box and click on OK to run the analysis.

SPSS Output 19.4 shows the main table for the model. Compare this table with SPSS Output 19.3 and you'll see that again there is no difference: we get a significant effect of surgery with an F of 4.08, $p < .05$, and a significant effect of baseline quality of life with an F of 217.25, $p < .001$. We can also see that the regression coefficient for surgery is –1.70. Again, the results are pretty similar to when we ran the analysis as ANCOVA (the values are

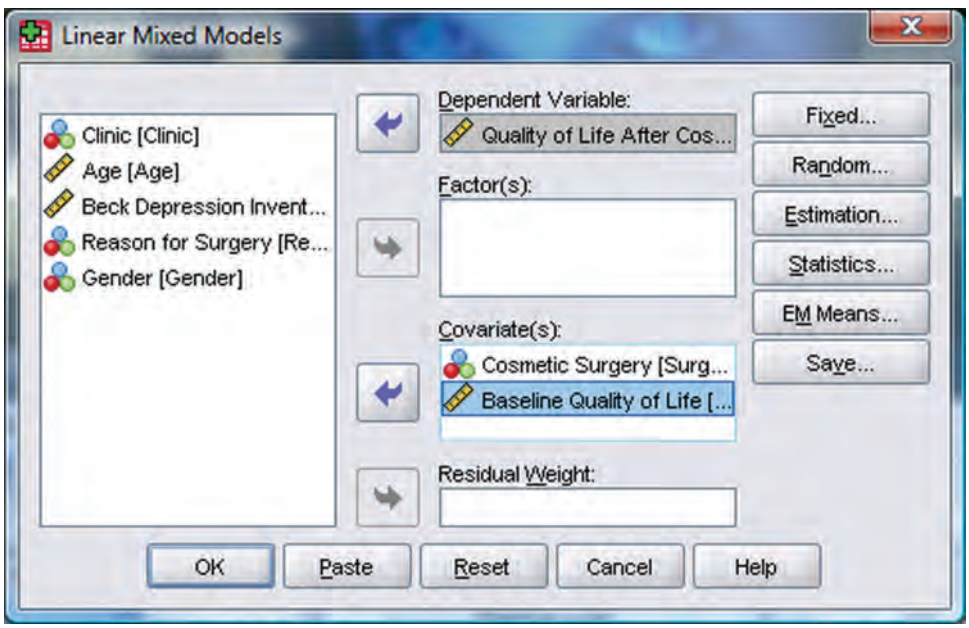
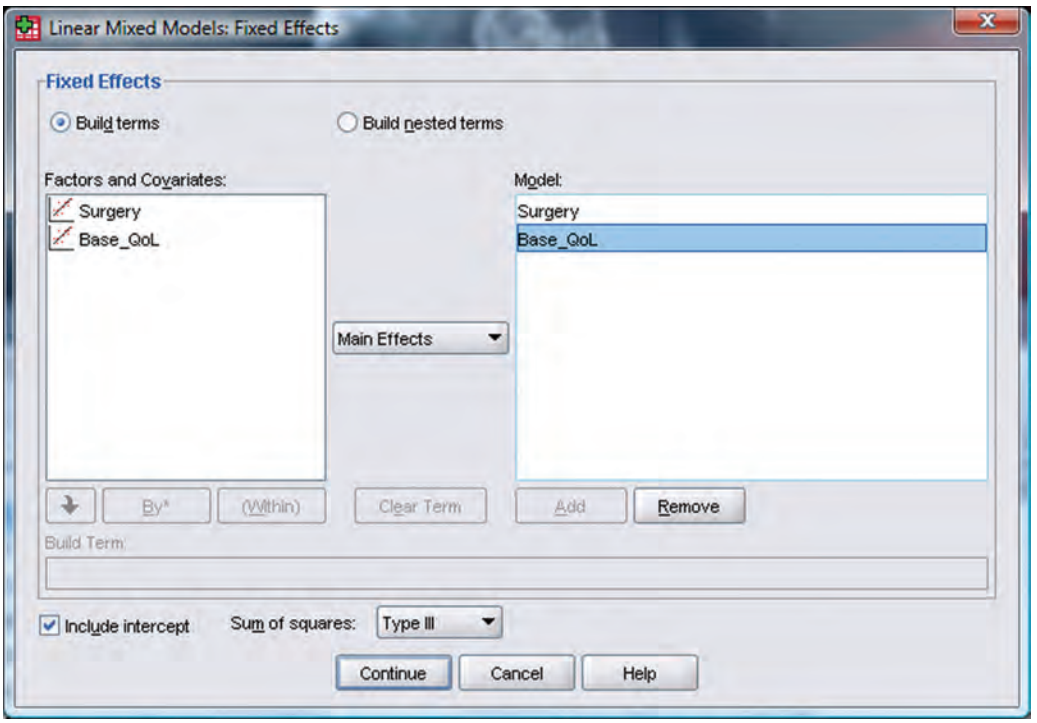


FIGURE 19.12 The main mixed models dialog box

FIGURE 19.13
The dialog box for specifying fixed effects in mixed models



SPSS OUTPUT 19.4

Model Dimension[a]

		Number of Levels	Number of Parameters
Fixed Effects	Intercept	1	1
	Surgery	1	1
	Base_QoL	1	1
	Residual		1
	Total	3	4

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Information Criteria[a]

-2 Log Likelihood	1852.543
Akaike's Information Criterion (AIC)	1860.543
Hurvich and Tsai's Criterion (AICC)	1860.690
Bozdogan's Criterion (CAIC)	1879.024
Schwarz's Bayesian Criterion (BIC)	1875.024

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	18.147025	2.891820	276.000	6.275	.000	12.454198	23.839851
Surgery	-1.697233	.839442	276	-2.022	.044	-3.349756	-.044710
Base_QoL	.665036	.045120	276.000	14.739	.000	.576213	.753858

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	276.000	39.379	.000
Surgery	1	276	4.088	.044
Base_QoL	1	276.000	217.249	.000

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

slightly different because here we're using maximum likelihood methods to estimate the parameters of the model but in ANCOVA we use ordinary least squares methods). Hopefully this has convinced you that we're just doing a regression here, something you have been doing throughout this book. This technique isn't radically different, and if you think about it as just an extension of what you already know, then it's relatively easy to understand. So, having shown you that we can do basic analyses through the mixed models command, let's now use its power to factor in the hierarchical structure of the data.

19.6.4. Factoring in the data structure: random intercepts ③

We have seen that when we factor in the pre-surgery quality of life scores, which themselves significantly predict post-surgery quality of life scores, surgery seems to positively affect quality of life. However, at this stage we have ignored the fact that our data have a hierarchical structure. Essentially we have violated the independence assumption because scores from people who had their surgery at the same clinic are likely to be related to each other (and certainly more related than with people at different clinics). We have seen that violating the assumption of independence can have some quite drastic consequences (see section 10.2.10). However, rather than just panic and gibber about our F-ratio being inaccurate, we can model this covariation within clinics explicitly by including the hierarchical data structure in our analysis.

To begin with, we will include the hierarchy in a fairly crude way by assuming simply that intercepts vary across clinics. Our model is now described by:

$$\text{QoL After Surgery}_{ij} = b_{0j} + b_1\text{Surgery}_{ij} + b_2\text{QoL Before Surgery}_{ij} + \varepsilon_{ij}$$
$$b_{0j} = b_0 + u_{0j}$$

To run a multilevel model we use the *Mixed Models* option by selecting Analyze Mixed Models Linear..., which will bring up the dialog box in Figure 19.8. This time we don't want to ignore this dialog box, but instead want to specify our level 2 variable (**Clinic**). We specify contextual variables that group participants (or subjects) in the box labelled *Subjects*. Select **Clinic** from the list of variables and drag it to the box labelled *Subjects* (or click on). The completed dialog box is shown in Figure 19.14.

Click on Continue to access the main dialog box. We don't need to change this because all we are doing in this model is changing the intercept from being fixed to random. Therefore, the main dialog box should still look like Figure 19.12. We also don't need to re-specify our fixed effects so there is no need to click on Fixed... unless you want to check that the dialog box still looks like Figure 19.13. However, we do need to specify a random effect for the first time, so click on Random... in the main dialog box to access the dialog box in Figure 19.15. The first thing we need to do is to specify our contextual variable. We do this by selecting it from the list of contextual variables that we have told SPSS about in Figure 19.14. These appear in the section labelled *Subjects* and because we only specified one variable, there is only one variable in the list, **Clinic**. Select this variable and drag it to the area labelled *Combinations* (or click on). We want to specify only that the intercept is random, and we do this by selecting Include intercept. Notice in this dialog box that there is a drop-down list to specify the type of covariance (Variance Components). For a random intercept model this default option is fine. Click on Continue to return to the main dialog box and then click on OK to run the analysis.

The output of this analysis is shown in SPSS Output 19.5. The first issue is whether allowing the intercepts to vary has made a difference to the model. We can test this from the change in the –2 log-likelihood (equation (19.10)). In our new model the –2LL is

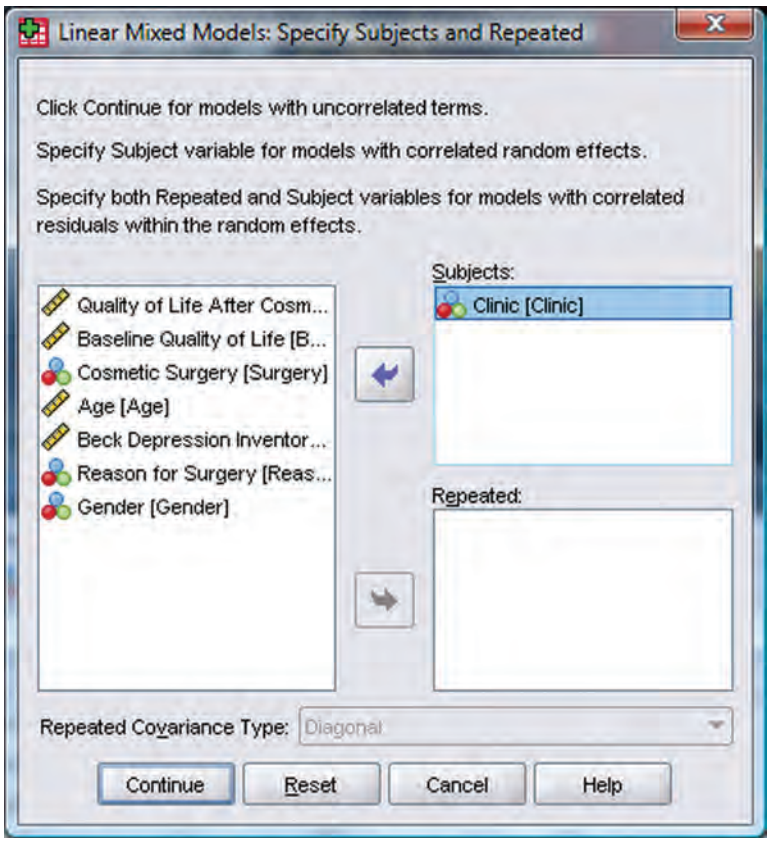


FIGURE 19.14
Specifying a level 2 variable in a hierarchical linear model

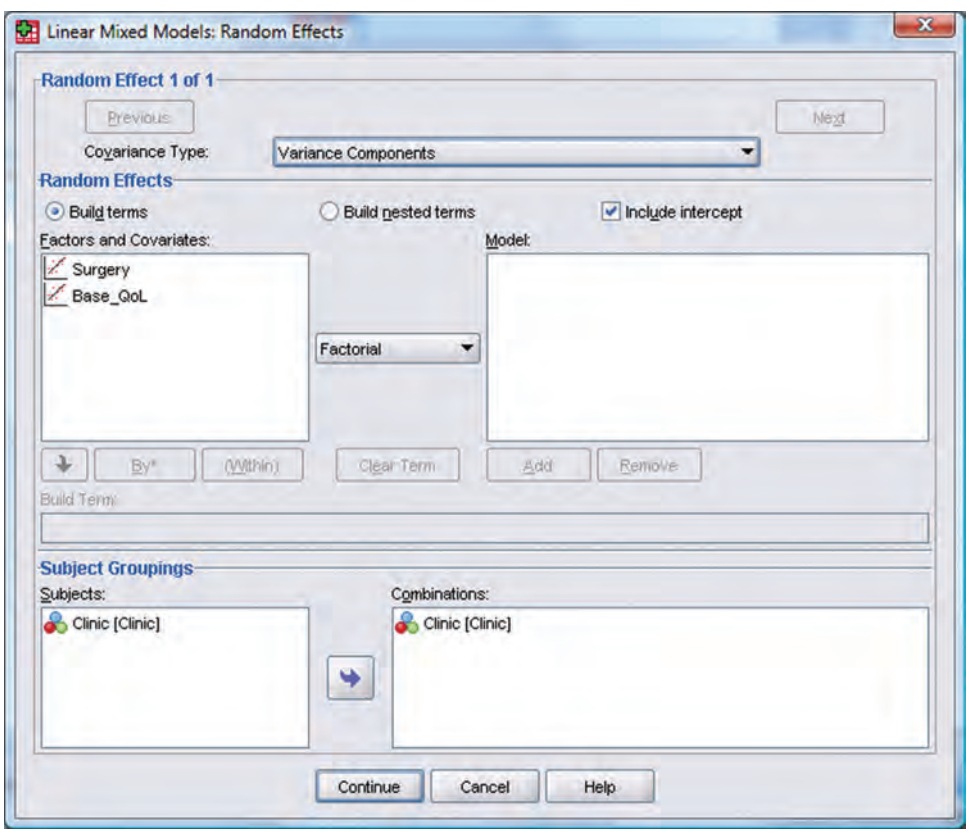


FIGURE 19.15
The dialog box for specifying random effects in mixed models

1837.49 (SPSS Output 19.5) based on a total of five parameters. In the old model (SPSS Output 19.4) the –2LL was 1852.54, based on four parameters. Therefore:

$$\chi^2_{\text{Change}} = 1852.54 - 1837.49 = 15.05$$
$$df_{\text{Change}} = 5 - 4 = 1$$

If we look at the critical values for the chi-square statistic with 1 degree of freedom in Appendix A4, they are 3.84 ($p < .05$) and 6.63 ($p < .01$); therefore, this change is highly significant. Put another way, it is important that we modelled this variability in intercepts because when we do our model is significantly improved. We can conclude then that the intercepts for the relationship between surgery and quality of life (when controlling for baseline quality of life) vary significantly across the different clinics.

SPSS OUTPUT 19.5

Model Dimension[b]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	Surgery	1		1	
	Base_QoL	1		1	
Random Effects	Intercept[a]	1	Variance Components	1	Clinic
Residual				1	
Total		4		5	

df for – 2LL

a. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using SPSS 11 syntax, please consult the current syntax reference guide for more information.

b. Dependent Variable: Quality of Life After Cosmetic Surgery.

Information Criteria[a]

-2 Log Likelihood	1837.490
Akaike's Information Criterion (AIC)	1847.490
Hurvich and Tsai's Criterion (AICC)	1847.712
Bozdogan's Criterion (CAIC)	1870.592
Schwarz's Bayesian Criterion (BIC)	1865.592

– 2LL

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	163.879	73.305	.000
Surgery	1	275.631	.139	.709
Base_QoL	1	245.020	83.159	.000

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

bs

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	29.563601	3.452958	163.879	8.562	.000	22.745578	36.381624
Surgery	-.312999	.838551	275.631	-.373	.709	-1.963776	1.337779
Base_QoL	.478630	.052486	245.020	9.119	.000	.375248	.582012

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$\text{Var}(\varepsilon_{ij})$

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		42.497179	3.703949	11.473	.000	35.823786	50.413718
Intercept [subject = Clinic]	Variance	9.237126	5.461678	1.691	.091	2.898965	29.432742

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$Var(u_{0j})$

You will also notice that the significance of the variance estimate for the intercept (9.24) is tested using the Wald statistic, which is a standard *z*-score in this case ($z = 1.69$). You should be cautious in interpreting the Wald statistic because, for random parameters especially, it can be quite unpredictable (for fixed effects it should be OK). The change in the –2LL is much more reliable, and you should use this to assess the significance of changes to the model – just like with logistic regression (Chapter 8).

By allowing the intercept to vary we also have a new regression parameter for the effect of surgery, which is –.31 compared to –1.70 when the intercept was fixed (SPSS Output 19.4). In other words, by allowing the intercepts to vary over clinics, the effect of surgery has decreased dramatically. In fact, it is not significant any more, $F(1, 275.63) = 0.14$, $p > .05$. This shows how, had we ignored the hierarchical structure in our data, we would have reached very different conclusions to what we have found here.

19.6.5. Factoring in the data structure: random intercepts and slopes ④

We have seen that including a random intercept is important for this model (it changes the log-likelihood significantly). However, we could now look at whether adding a random slope will also be beneficial by adding this term to the model. The model is now described by equation (19.9), which we saw earlier on; it can be specified in SPSS with only minor modifications to the dialog boxes. All we are doing is adding another random term to the model; therefore, the only changes we need to make are in the dialog box accessed by clicking on Random... . (If you are starting from scratch then follow the instructions for setting up the dialog box in the previous section.) We need to select the predictor (**Surgery**) from the list of *Factors and covariates* and add it to the model by clicking on Add (see Figure 19.16). Click on Continue to return to the main dialog box and then click on OK to run the analysis

All we're interested in at this stage is estimating the effect of including the variance in intercepts. SPSS Output 19.6 gives us the –2LL for the new model and the value of the variance in slopes (29.63). To find the significance of the variance in slopes, we subtract this value from the –2LL for the previous model. This gives us a chi-square statistic with $df = 1$ (because we have added only one new parameter to the model: the variance in slopes). In our new model the –2LL is 1816 (SPSS Output 19.6) based on a total of six parameters. In the old model (SPSS Output 19.5) the –2LL was 1837.49, based on five parameters. Therefore:

$$\chi^2_{\text{Change}} = 1837.49 - 1816 = 21.49$$
$$df_{\text{Change}} = 6 - 5 = 1$$

Comparing this value to the same critical values as before for the chi-square statistic with $df = 1$ (i.e. 3.84 and 6.63) shows that this change is highly significant because 21.49 is much larger than these two values. Put another way, the fit of our model significantly improved when the variance of slopes was included: there is significant variability in slopes.

Now that we know that there is significant variability in slopes, we can look to see whether the slopes and intercepts are correlated (or covary). By selecting Variance Components in the previous analysis, we assumed that the covariances between the intercepts and slopes were zero. Therefore, SPSS estimated only the variance of slopes. This was a useful thing

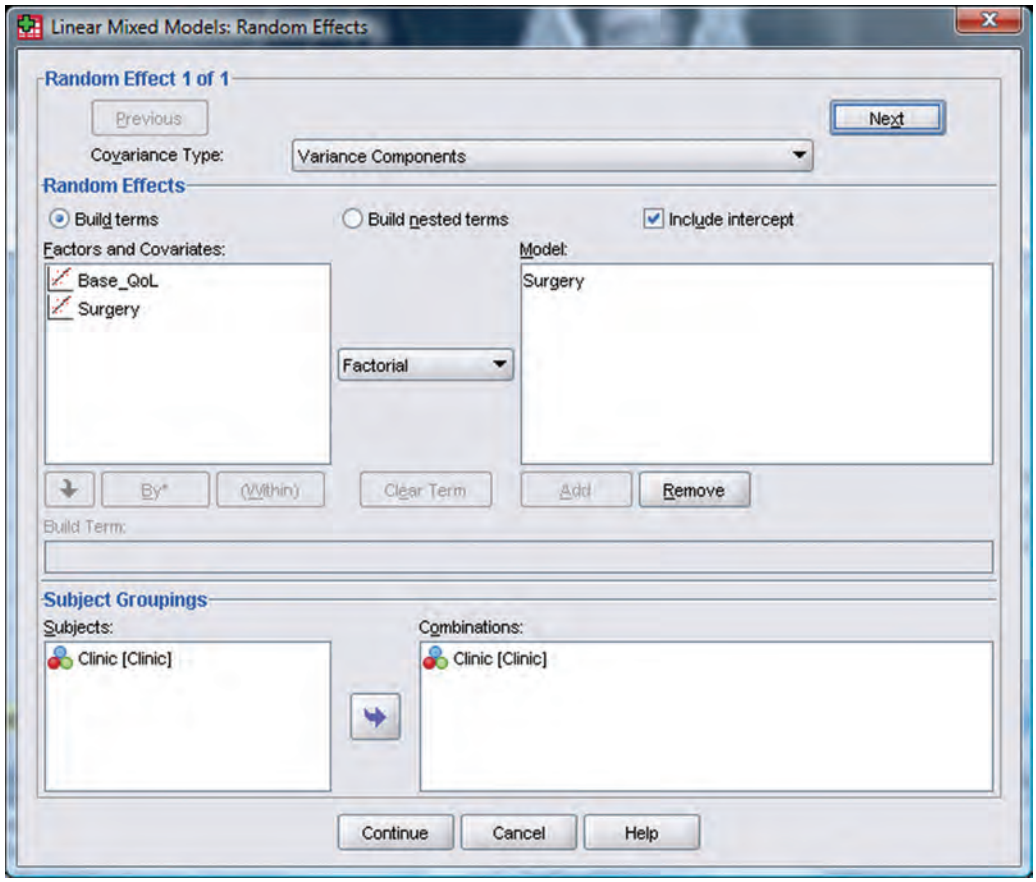


FIGURE 19.16 The dialog box for specifying random effects in mixed models

to do because it allowed us to look at the effect of the variance of slopes in isolation. If we now want to include the covariance between random slopes and random intercepts we do this by clicking on Variance Components in Figure 19.16 to access the drop-down list, and selecting Unstructured instead. By changing to Unstructured, we remove the assumption that the covariances between slopes and intercepts are zero, and so SPSS will estimate this covariance. As such, by changing to Unstructured, we add a new term to the model that estimates the covariance between random slopes and intercepts. Redo the analysis but change Variance Components to Unstructured in Figure 19.16.

The output of this analysis is shown in SPSS Output 19.7. The first issue is whether adding the covariance between slopes and intercepts has made a difference to the model using the change in the –2LL (equation (19.10)). In our new model the –2LL is 1798.62 (SPSS Output 19.7) based on a total of seven parameters. In the old model (SPSS Output 19.6) the –2LL was 1816, based on six parameters. Therefore:

$$\chi^2_{\text{Change}} = 1816 - 1798.62 = 17.38$$
$$df_{\text{Change}} = 7 - 6 = 1$$

This change is highly significant at $p < .01$ because 17.38 is bigger than the critical value of 6.63 for the chi-square statistic with 1 degree of freedom (see Appendix A4). Put another way, our model is significantly improved when the covariance term is included in

SPSS OUTPUT 19.6

Information Criteria[a]

-2 Log Likelihood	1816.001
Akaike's Information Criterion (AIC)	1828.001
Hurvich and Tsai's Criterion (AICC)	1828.314
Bozdogan's Criterion (CAIC)	1855.724
Schwarz's Bayesian Criterion (BIC)	1849.724

–2LL

The information criteria are displayed in smaller-is-better forms.

. Dependent Variable: Quality of Life After Cosmetic Surgery.

$\text{Var}(\varepsilon_{ij})$

$\text{Var}(u_{0j})$

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		35.008422	3.132866	11.175	.000	29.376457	41.720130
Intercept [subject = Clinic]	Variance	33.181911	16.900824	1.963	.050	12.227895	90.043233
Surgery [subject = Clinic]	Variance	29.630281	16.497840	1.796	.072	9.949366	88.242166

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$\text{Var}(u_{1j})$

the model. The variance estimates for the intercept (37.60) and slopes (–36.68 and 38.41), and their associated significance based on the Wald test, confirm this because all three estimates are close to significance (although I reiterate my earlier point that the Wald statistic should be interpreted with caution).

One important part of the output to take note of is that the random part of the slopes now has two values (–36.68 and 38.41). The reason that there are two values is because we changed from a covariance structure of Variance Components, which assumes that parameters are uncorrelated to Unstructured, which makes no such assumption, and, therefore, the covariance is estimated too. The first of these values is the covariance between the random slope and random intercept, and the second is the variance of the random slopes. We encountered covariance in Chapter 6 and saw that it is an unstandardized measure of the relationship between variables. In other words, it's like a correlation. Therefore, the covariance term tells us whether there is a relationship or interaction between the random slope and the random intercept within the model. The actual size of this value is not terribly important because it is unstandardized (so we can't compare the size of covariances measured across different variables), but the direction of it is. In this case the covariance is negative (–36.68) indicating a negative relationship between the intercepts and the slopes. Remember that we are looking at the effect of surgery on quality of life in 10 different clinics, so this means that, across these clinics, as the intercept for the relationship between surgery and quality of life increases, the value of the slope decreases. This is best understood using a diagram and Figure 19.17 shows the observed values of quality of life after surgery plotted against those predicted by our model. In this diagram each line represents a different clinic. We can see that the 10 clinics differ: those with low intercepts (low values on the *y*-axis) have quite steep positive slopes. However, as the intercept increases (as we go from the line that crosses the *y*-axis at the lowest point up to the line that hits the *y*-axis at the highest point) the slopes of the lines get flatter (the slope decreases). The negative covariance between slope and intercept reflects this relationship. Had it been positive it would mean the opposite: as intercepts increase, the slopes increase also.

The second term that we get with the random slope is its variance (in this case 38.41). This tells us how much the slopes vary around a single slope fitted to the entire data set (i.e. ignoring the clinic from which the data came). This confirms what our chi-square test showed us: that the slopes across clinics are significantly different.

SPSS OUTPUT 19.7

Model Dimension[b]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	Base_QoL	1		1	
	Surgery	1		1	
Random Effects	Intercept + Surgery[a]	2	Unstructured	3	Clinic
Residual				1	
Total		5		7	

df for –2LL

a. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using SPSS 11 syntax, please consult the current syntax reference guide for more information.

b. Dependent Variable: Quality of Life After Cosmetic Surgery.

Information Criteria[a]

-2 Log Likelihood	1798.624
Akaike's Information Criterion (AIC)	1812.624
Hurvich and Tsai's Criterion (AICC)	1813.042
Bozdogan's Criterion (CAIC)	1844.967
Schwarz's Bayesian Criterion (BIC)	1837.967

–2LL

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	84.954	107.284	.000
Surgery	1	9.518	.097	.762
Base_QoL	1	265.933	33.984	.000

bs

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	40.102525	3.871729	84.954	10.358	.000	32.404430	47.800620
Surgery	-.654530	2.099413	9.518	-.312	.762	-5.364643	4.055583
Base_QoL	.310218	.053214	265.933	5.830	.000	.205443	.414993

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$Var(u_{0j})$ = Variance of intercepts

$Var(\varepsilon_{ij})$ = Variance of residuals

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		34.955705	3.116670	11.216	.000	29.351106	41.630504
Intercept + Surgery [subject = Clinic]	UN (1,1)	37.609439	18.726052	2.008	.045	14.173482	99.796926
	UN (2,1)	-36.680707	18.763953	-1.955	.051	-73.457378	.095965
	UN (2,2)	38.408857	20.209811	1.901	.057	13.694612	107.724141

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$Cov(u_{0j}, u_{1j})$ = Covariance between intercepts and Slopes

$Var(u_{1j})$ = Variance of Slopes

We can conclude then that the intercepts and slopes for the relationship between surgery and quality of life (when controlling for baseline quality of life) vary significantly across the different clinics. By allowing the intercept and slopes to vary we also have a new regression parameter for the effect of surgery, which is –.65 compared to –0.31 when the slopes were fixed (SPSS Output 19.5). In other words, by allowing the intercepts to vary over clinics, the effect of surgery has increased slightly, although it is still nowhere near significant, $F(1, 9.518) = 0.10$, $p > .05$. This shows how, had we ignored the hierarchical structure in our data, we would have reached very different conclusions to what we have found here.

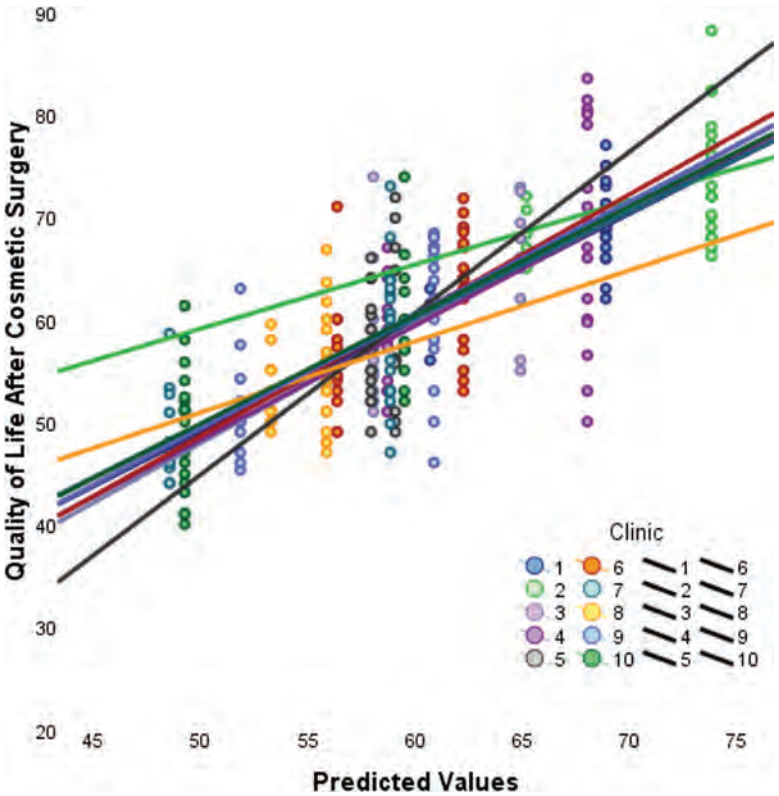


FIGURE 19.17 Predicted values from the model (surgery predicting quality of life after controlling for baseline quality of life) plotted against the observed values

19.6.6. Adding an interaction to the model ④

We can now build up the model by adding in another variable. One of the variables we measured was the reason for the person having cosmetic surgery: was it to resolve a physical problem or was it purely for vanity? We can add this variable to the model, and also look at whether it interacts with surgery in predicting quality of life.[5] Our model will simply expand to incorporate these new terms, and each term will have a regression coefficient (which we select to be fixed). Therefore, our new model can be described as in the equation below (note that all that has changed is that there are two new predictors):

$$\begin{aligned} \text{QoL After}_{ij} &= b_{0j} + b_{1j}\text{Surgery}_{ij} + b_2\text{QoL Before Surgery}_{ij} + b_3\text{Reason}_{ij} \\ &\quad + b_4\ (\text{Reason} \times \text{Surgery})_{ij} + \varepsilon_{ij} \\ b_{0j} &= b_0 + u_{0j} \\ b_{1j} &= b_1 + u_{1j} \end{aligned} \tag{19.11}$$

To set up this model in SPSS is very easy to do and just requires some minor changes to the dialog boxes that we have already used. First, select Analyze Mixed Models Linear...; this initial dialog box should be set up as for the previous analysis, but if you're running this analysis without running the prior one first then set up the level 2 variable of

[5] In reality, because we would use the change in the –2LL to see whether effects are significant, we would build this new model up a term at a time. Therefore, we would first include only **Reason** in the model, then in a separate analysis we would add the interaction. By doing so we can calculate the change in –2LL for each effect. To save space I'm going to put both into the model in a single step.

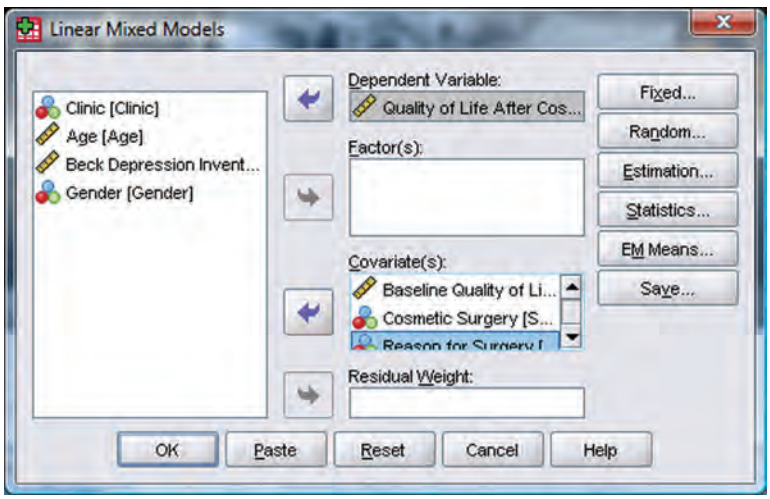


FIGURE 19.18 The main mixed models dialog box

Clinic as in the previous sections (the completed dialog box is shown in Figure 19.14). Click on Continue to access the main dialog box. If you're continuing the previous analysis then this dialog box will already be set up with the previous model. If you've jumped straight into this analysis then set this dialog box up as with our previous model (Figure 19.12). We have two new covariates to add to the model: the effect of the reason for the surgery (**Reason**) and the interaction of **Reason** and **Surgery**. At this stage we simply need to add **Reason** as a covariate, so select this variable and drag it to the space labelled *Covariate(s)* (or click on). The completed dialog box is in Figure 19.18.

We need to add these fixed effects to our model, so click on Fixed... to bring up the dialog box in Figure 19.19. First let's specify the main effect of **Reason**; to do this, select this variable in the list labelled *Factors and Covariates* and then click on Add to transfer it to *Model*. To specify the interaction term, first click on Factorial and change it to Interaction. Next, select **Surgery** from *Factors and Covariates* and then while holding down the *Ctrl* key select **Reason**. With both variables selected click on Add to transfer them to *Model* as an interaction effect. The dialog box should now look like Figure 19.19. Click on Continue to return to the main dialog box. We don't need to specify any extra random coefficients so we can leave the dialog box accessed through Random... as it is in Figure 19.16, and we can leave the other options as they are in previous analyses. In the main dialog box click on OK to run the analysis.

SPSS Output 19.8 shows the resulting output, which is similar to the previous output except that we now have two new fixed effects. The first issue is whether these new effects make a difference to the model. We can use the log-likelihood statistics again:

$$\chi^2_{\text{Change}} = 1798.62 - 1789.05 = 9.57$$
$$df_{\text{Change}} = 9 - 7 = 2$$

If we look at the critical values for the chi-square statistic in the Appendix, it is 5.99 ($p < .05$, $df = 2$); therefore, this change is significant. We can look at the effects individually in the table of fixed effects. This tells us that quality of life before surgery significantly predicted quality of life after surgery, $F(1, 268.92) = 33.65$, $p < .001$, surgery still did not significantly predict quality of life, $F(1, 15.86) = 2.17$, $p = .161$, but the reason for surgery, $F(1, 259.89) = 9.67$, $p < .01$, and the interaction of the reason for surgery and surgery, $F(1, 217.09) = 6.28$, $p < .05$, both did significantly predict quality of life. The

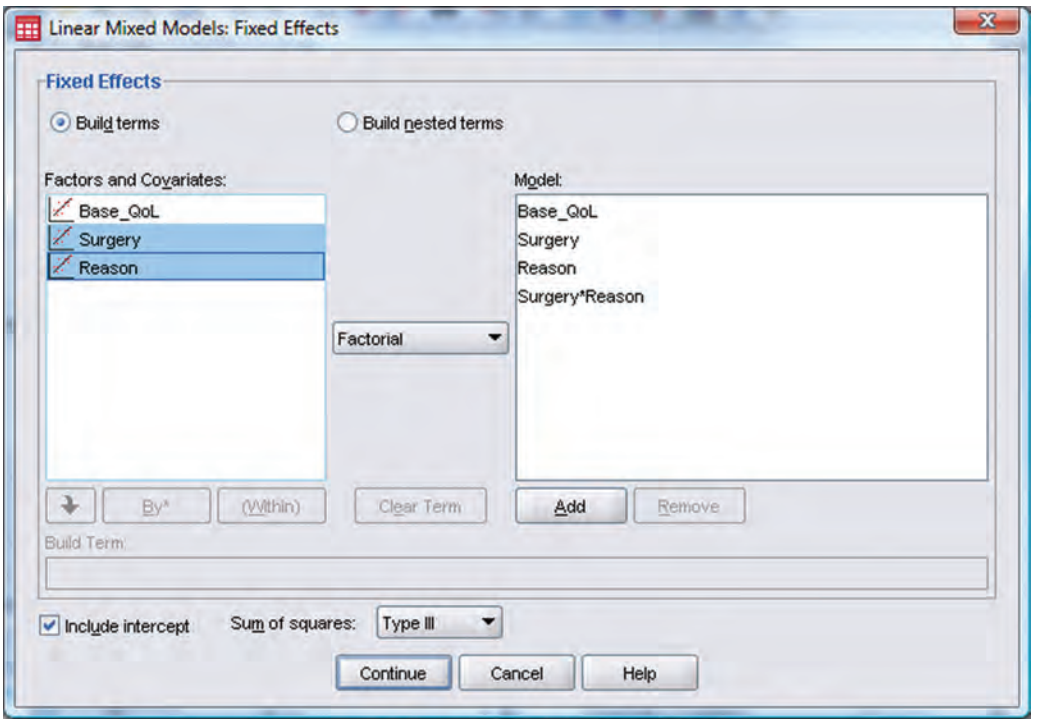


FIGURE 19.19 Specifying a fixed effect interaction in mixed models

table of estimates of fixed effects tells us much the same thing except it also gives us the regression coefficients and their confidence intervals.

The values of the variance for the intercept (30.06) and the slope (29.35) are lower than the previous model but still significant (one-tailed). Also the covariance between the slopes and intercepts is still negative (–28.08). As such our conclusions about our random parameters stay much the same as in the previous model.

The effect of the reason for surgery is easy to interpret. Given that we coded this predictor as 1 = physical reason and 0 = change appearance, the negative coefficient tells us that as reason increases (i.e. for surgery = 0, as a person goes from changing their appearance to a physical reason) quality of life decreases. However, this effect in isolation isn't that interesting because it includes both people who had surgery and the waiting list controls. More interesting is the interaction term, because this takes account of whether or not the person had surgery. To break down this interaction we could rerun the analysis separately for the two 'reason groups'. Obviously we would remove the interaction term and the main effect of **Reason** from this analysis (because we are analysing the physical reason group separately from the group that wanted to change their appearance). As such, you need to fit the model in the previous section, but first split the file by **Reason**.

SELF-TEST Split the file by **Reason** and then run a multilevel model predicting **Post_QoL** with a random intercept, and random slopes for **Surgery**, and including **Base_QoL** and **Surgery** as predictors.

SPSS OUTPUT 19.8

Model Dimension[b]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	Base_QoL	1		1	
	Surgery	1		1	
	Reason	1		1	
	Surgery * Reason	1		1	
Random Effects	Intercept + Surgery[a]	2	Unstructured	3	Clinic
Residual				1	
Total		7		9	

df for –2LL

a. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using SPSS 11 syntax, please consult the current syntax reference guide for more information.

b. Dependent Variable: Quality of Life After Cosmetic Surgery.

Information Criteria[a]

-2 Log Likelihood	1789.045
Akaike's Information Criterion (AIC)	1807.045
Hurvich and Tsai's Criterion (AICC)	1807.722
Bozdogan's Criterion (CAIC)	1848.629
Schwarz's Bayesian Criterion (BIC)	1839.629

–2LL

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	108.853	122.593	.000
Base_QoL	1	268.920	33.647	.000
Surgery	1	15.863	2.167	.161
Reason	1	259.894	9.667	.002
Surgery * Reason	1	217.087	6.278	.013

bs

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	42.517820	3.840055	108.853	11.072	.000	34.906839	50.128800
Base_QoL	.305356	.052642	268.920	5.801	.000	.201713	.408999
Surgery	-3.187677	2.165484	15.863	-1.472	.161	-7.781510	1.406157
Reason	-3.515148	1.130552	259.894	-3.109	.002	-5.741357	-1.288939
Surgery * Reason	4.221288	1.684798	217.087	2.506	.013	.900633	7.541944

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$Var(u_{0j})$ = Variance of intercepts

$Var(\varepsilon_{ij})$ = Variance of residuals

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		33.859719	3.024395	11.196	.000	28.421886	40.337948
Intercept + Surgery [subject = Clinic]	UN (1,1)	30.056340	15.444593	1.946	.052	10.978478	82.286775
	UN (2,1)	-28.083657	15.195713	-1.848	.065	-57.866706	1.699393
	UN (2,2)	29.349323	16.404492	1.789	.074	9.813593	87.774453

a. Dependent Variable: Quality of Life After Cosmetic Surgery.

$Cov(u_{0j}, u_{1j})$ = Covariance between intercepts and Slopes

$Var(u_{1j})$ = Variance of Slopes

SPSS Output 19.9 shows the parameter estimates from these analyses. It shows that for those operated on only to change their appearance, surgery almost significantly predicted quality of life after surgery, $b = -4.31$, $t(7.72) = -1.92$, $p = .09$. The negative gradient shows that in these people, quality of life was lower after surgery compared to the control group. However, for those that had surgery to solve a physical problem surgery did not significantly predict quality of life, $b = 1.20$, $t(7.61) = 0.58$, $p = .58$. However, the slope was positive indicating that people who had surgery scored higher on quality of life than

SPSS OUTPUT 19.9 Surgery to Change Appearance:

Estimates of Fixed Effects[a,b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	41.786055	5.487873	77.331	7.614	.000	30.859052	52.713059
Base_QoL	.338492	.079035	88.619	4.283	.000	.181440	.495543
Surgery	-4.307014	2.239912	7.719	-1.923	.092	-9.505157	.891130

a. Reason for Surgery = Change Appearance

b. Dependent Variable: Quality of Life After Cosmetic Surgery.

Surgery for a Physical Problem:

Estimates of Fixed Effects[a,b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	38.020790	4.666154	93.558	8.148	.000	28.755460	47.286119
Base_QoL	.317710	.068883	172.816	4.612	.000	.181749	.453670
Surgery	1.196550	2.081999	7.614	.575	.582	-3.647282	6.040382

a. Reason for Surgery = Physical reason

b. Dependent Variable: Quality of Life After Cosmetic Surgery.

those on the waiting list (although not significantly so!). The interaction effect, therefore, reflects the difference in slopes for surgery as a predictor of quality of life in those that had surgery for physical problems (slight positive slope) and those that had surgery purely for vanity (a negative slope).

We could sum up these results by saying that quality of life after surgery, after controlling for quality of life before surgery, was lower for those that had surgery to change their appearance than those that had surgery for a physical reason. This makes sense because for those having surgery to correct a physical problem, the surgery has probably bought relief and so their quality of life will improve. However, for those having surgery for vanity they might well discover that having a different appearance wasn't actually at the root of their unhappiness, so their quality of life is lower.

CRAMMING SAM'S TIPS Multilevel models SPSS output

- The **Information Criteria** table can be used to assess the overall fit of the model. The –2LL can be significance tested with *df* = the number of parameters being estimated. It is mainly used, though, to compare models that are the same in all but one parameter by testing the difference in –2LL in the two models against df = 1 (if only one parameter has been changed). The AIC, AICC, CAIC and BIC can also be compared across models (but not significance tested).
- The table of **Type III Tests of Fixed Effects** tells you whether your predictors significantly predict the outcome: look in the column labelled *Sig.* If the value is less than .05 then the effect is significant.
- The table of **Estimates of Fixed Effects** gives us the regression coefficient for each effect and its confidence interval. The direction of these coefficients tells us whether the relationship between each predictor and the outcome is positive or negative.
- The table labelled **Estimates of Covariance Parameters** tells us about any random effects in the model. These values can tell us how much intercepts and slopes varied over our level 1 variable. The significance of these estimates should be treated cautiously. The exact labelling of these effects depends on which covariance structure you selected for the analysis.

19.7. Growth models ④

Growth models are extremely important in many areas of science including psychology, medicine, physics, chemistry or economics. In a growth model the aim is to look at the rate of change of a variable over time: for example, we could look at white blood cell counts, attitudes, radioactive decay or profits. In all cases we're trying to see which model best describes the change over time.

19.7.1. Growth curves (polynomials) ④

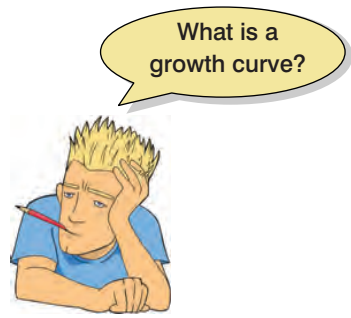


Figure 19.20 gives some examples of possible **growth curves**. This diagram shows three **polynomials** representing a linear trend (the blue line) otherwise known as a first-order polynomial, a quadratic trend (the green line) otherwise known as a second-order polynomial, and a cubic trend (the red line) otherwise known as a third-order polynomial. Notice first that the linear trend is a straight line, but as the polynomials increase they get more and more curved, indicating more rapid growth over time. Also, as polynomials increase, the change in the curve is quite dramatic (so dramatic that I adjusted the scale of the graph to fit all three curves on the same diagram). This observation highlights the fact that any growth curve higher than a quadratic (or possibly cubic) trend is very unrealistic in real data. By fitting a growth model to the data we can see which trend best describes the growth of an outcome variable over time (although, no one will believe that a significant fifth-order polynomial is telling us anything meaningful about the real world!).

The growth curves that we have described might seem familiar to you: they are the same as the trends that we described for ordered means in section 10.2.11.5. What we are discussing now is really no different. There are just two important things to remember when fitting growth curves: (1) you can fit polynomials up to one less than the number of time points that you have; and (2) a polynomial is defined by a simple power function. On the first point, this means that with three time points you can fit a linear and quadratic growth curve (or a first- and second-order polynomial), but you cannot fit any higher-order growth curves. Similarly, if you have six time points you can fit up to a fifth-order polynomial. This is the same basic idea as having one less contrast than the number of groups in ANOVA (see section 10.2.11).

On the second point, we have to define growth curves manually in multilevel models in SPSS: there is not a convenient option that we can select to do it for us. However, this is quite easy to do. If *time* is our predictor variable, then a linear trend is tested by including this variable alone. A quadratic or second-order polynomial is tested by including a predictor that is $time^2$, a cubic or third-order polynomial is tested by including a predictor that is $time^3$ and so on. So any polynomial is tested by including a variable that is the predictor to the power of the order of polynomial that you want to test: for a fifth-order polynomial we need a predictor of $time^5$ and for an n-order polynomial we would have to include $time^n$ as a predictor. Hopefully you get the general idea.

19.7.2. An example: the honeymoon period ②

I recently heard a brilliant talk given by Professor Daniel Kahneman, who won the 2002 Nobel Prize for Economics. In this talk Kahneman brought together an enormous amount of research on life satisfaction (he explored questions such as whether people are happier if they are richer). There was one graph in this talk that particularly grabbed my attention. It showed

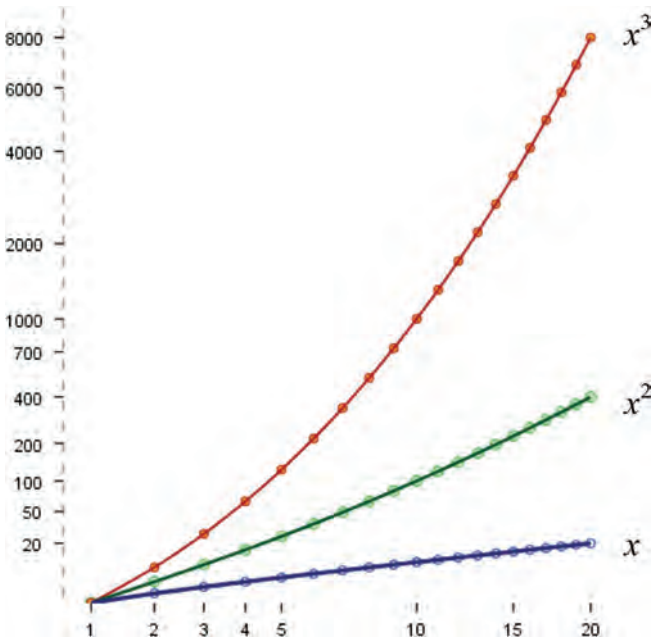


FIGURE 19.20
Illustration of a first-order (linear, blue), second-order (quadratic, green) and third-order (cubic, red) polynomial

that leading up to marriage people reported greater life satisfaction, but by about two years after marriage this life satisfaction decreased back to its baseline level. This graph perfectly illustrated what people talk about as the 'honeymoon period': a new relationship/marriage is great at first (no matter how ill suited you may be) but after six months or so the cracks start to appear and everything turns to elephant dung. Kahneman argued that people adapt to marriage; it does not make them happier in the long run (Kahneman & Krueger, 2006).[6] This got me thinking about relationships not involving marriage (is it marriage that makes you happy, or just being in a long-term relationship?). Therefore, in a completely fictitious parallel world where I don't research child anxiety, but instead concern myself with people's life satisfaction, I collected some data. I organized a massive speed-dating event (see Chapter 14). At the start of the night I measured everyone's life satisfaction (**Satisfaction_Baseline**) on a 10-point scale (0 = completely dissatisfied, 10 = completely satisfied) and their gender (**Gender**). After the speed dating I noted all of the people who had found dates. If they ended up in a relationship with the person that they met on the speed-dating night then I stalked these people over the next 18 months of that relationship. As such, I had measures of their life satisfaction at 6 months (**Satisfaction_6_Months**), 12 months (**Satisfaction_12_Months**) and 18 months (**Satisfaction_18_Months**), after they entered the relationship. None of the people measured were in the same relationship (i.e. I measured only life satisfaction from one of the people in the couple).[7] Also, as is often the case with longitudinal data, I didn't have scores for all people at all time points because not everyone was available at the follow-up sessions. One of the benefits of a multilevel approach is that these missing data do not pose a particular problem. The data are in the file **Honeymoon Period.sav**.

Figure 19.21 shows the data. Each circle is a data point and the line shows the average life satisfaction over time. Basically, from baseline, life satisfaction rises slightly at time 2 (6 months) but then starts to decrease over the next 12 months. There are two things to note about the data. First, time 0 is before the people enter into their new relationship yet

[6] The romantics among you might be relieved to know that others have used the same data to argue the complete opposite: that married people are happier than non-married people in the long term (Easterlin, 2003).

[7] However, I could have measured both people in the couple because using a multilevel model I could have treated people as being nested within 'couples' to take account of the dependency in their data.

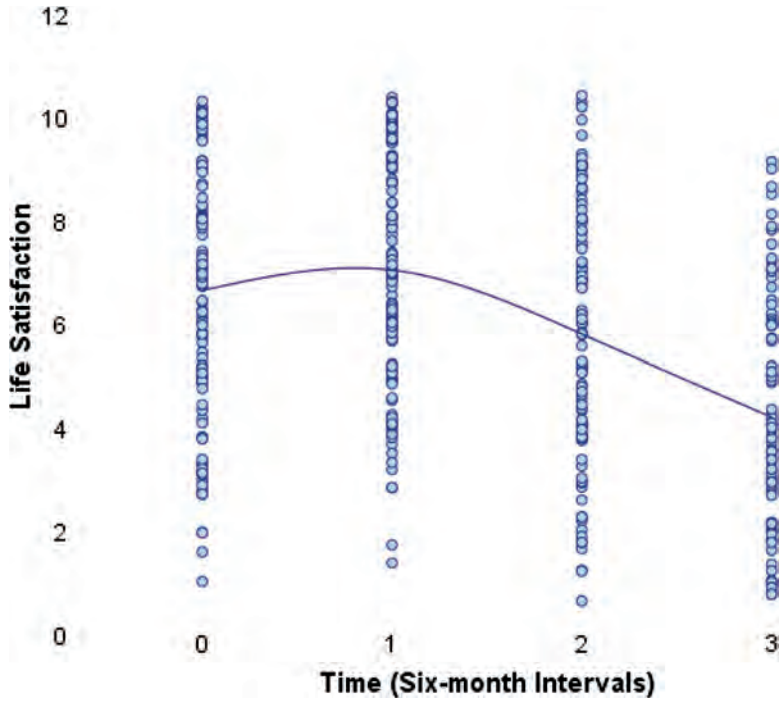
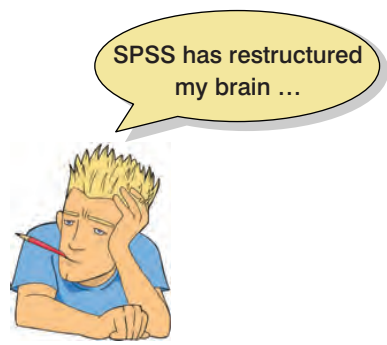


FIGURE 19.21
Life satisfaction over time

already there is a lot of variability in their responses (reflecting the fact that people will vary in their satisfaction due to other reasons such as finances, personality and so on). This suggests that intercepts for life satisfaction differ across people. Second, there is also a lot of variability in life satisfaction after the relationship has started (time 1) and at all subsequent time points, which suggests that the slope of the relationship between time and life satisfaction might vary across people also. If we think of the time points as a level 1 variable that is nested with people (a level 2 variable) then we can easily model this variability in intercepts and slopes within people. We have a situation similar to Figure 19.4 (except with two levels instead of three, although we could add in the location of the speed dating event as a level 3 variable if we had that information!).

19.7.3. Restructuring the data ③



The first problem with having data measured over time is that to do a multilevel model the data need to be in a different format to what we are used to. Figure 19.22 shows how we would normally set up the data editor for a repeated-measures design: each row represents a person, and notice that the repeated-measures variable of time is represented by four different columns. If we were going to run an ordinary repeated-measures ANOVA this data layout would be fine; however, for a multilevel model we need the variable **Time** to be represented by a single column. We could enter all of the data again, but that would be a pain; luckily we don't have to do this because SPSS has a *restructure* command, which is also a pain, but not as much as retyping the data. This command enables you to take your data set and create a new data set that is organized differently.

To access the restructure wizard select Data > Restructure... . The steps in the wizard are shown in Figure 19.23. In the first dialog box you need to say whether you are converting variables to

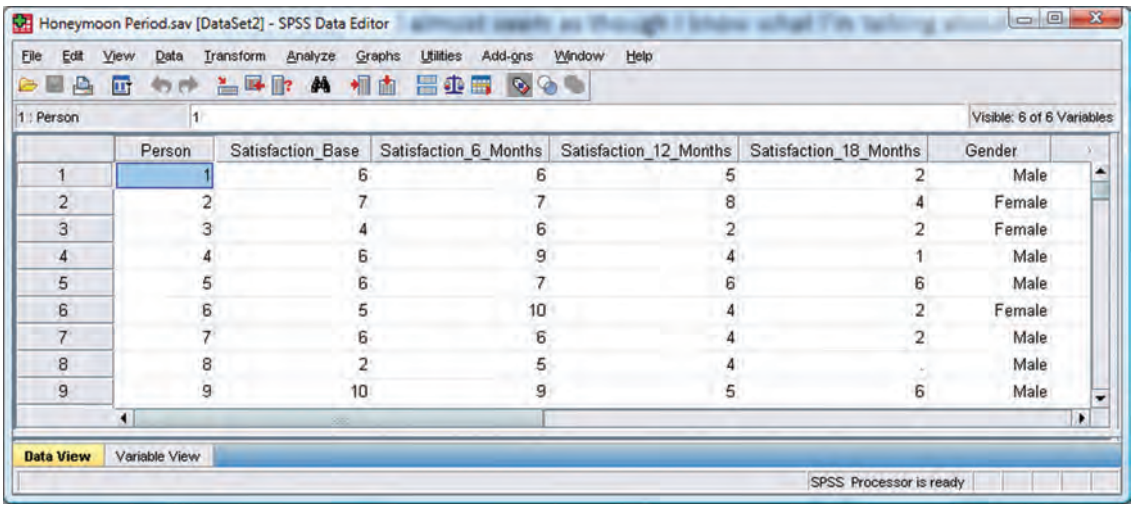
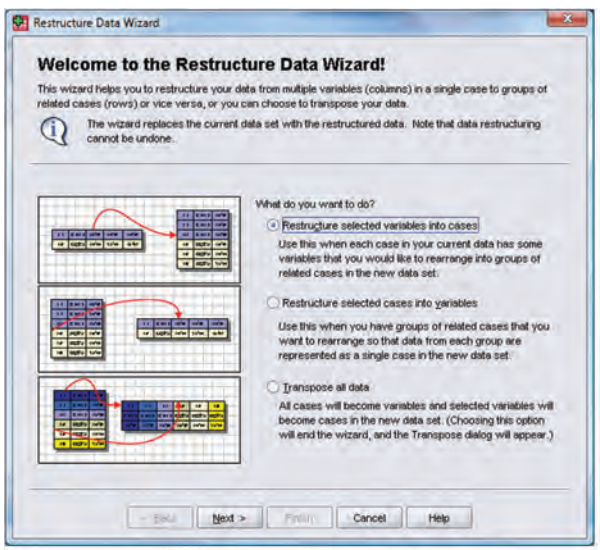
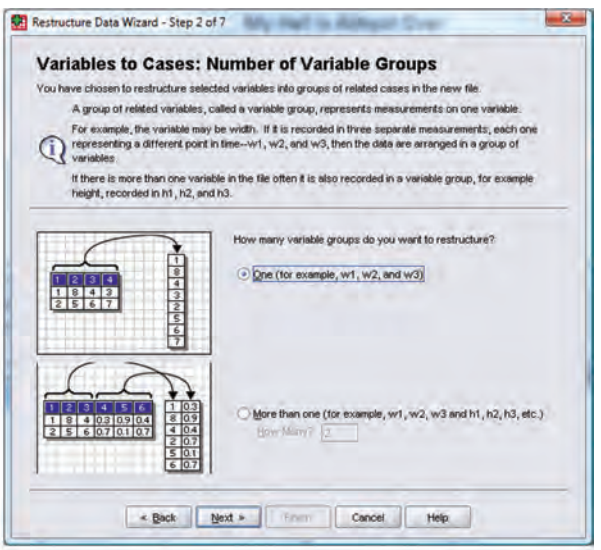
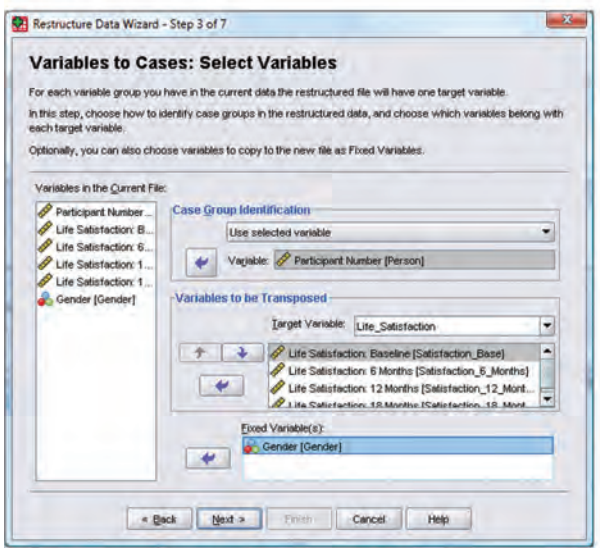


FIGURE 19.22 The data editor for a normal repeated-measures data set







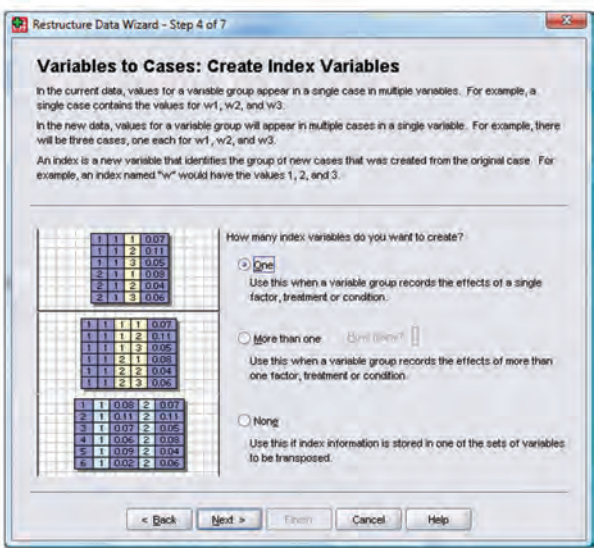


FIGURE 19.23 *Continued*

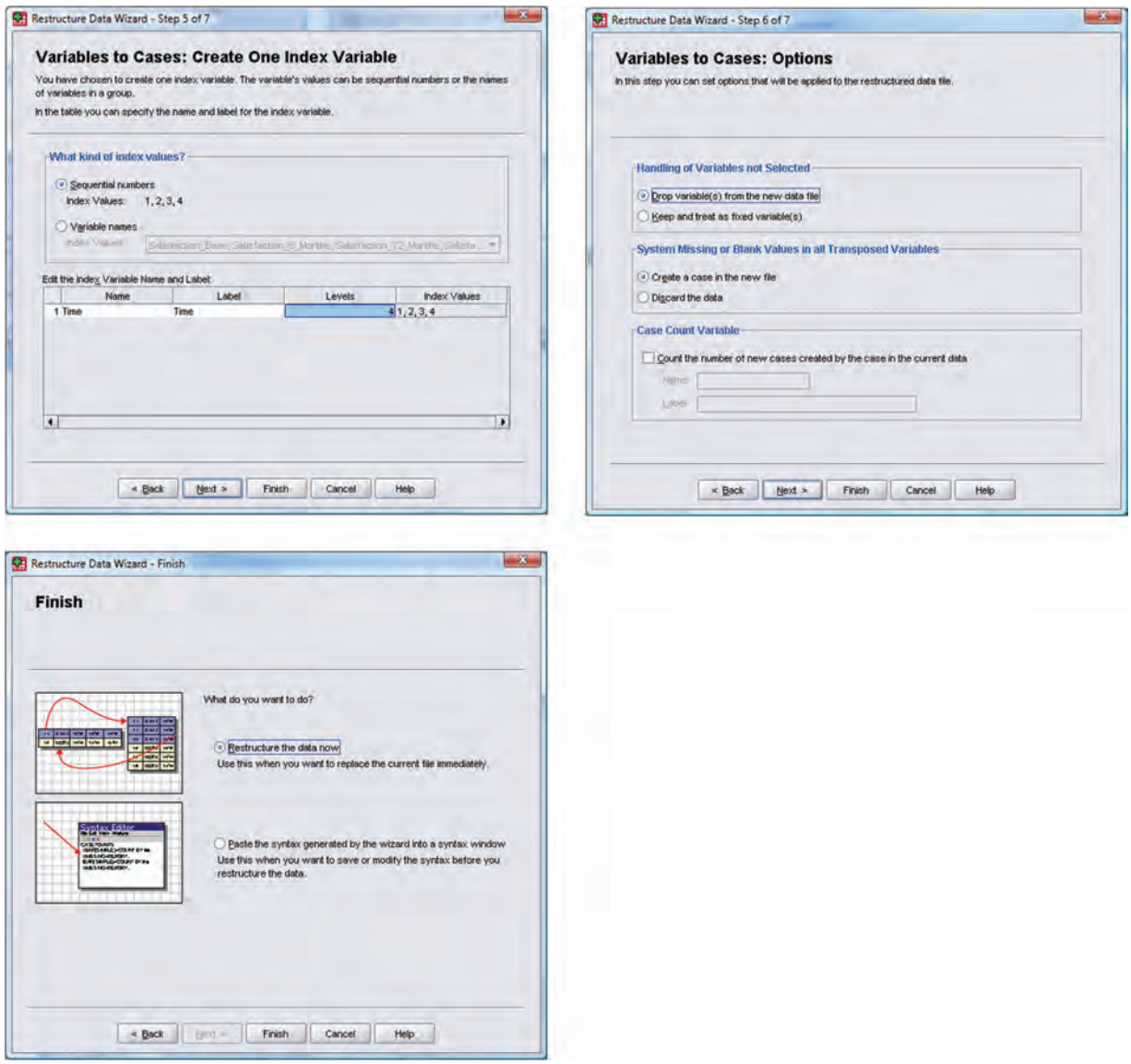


FIGURE 19.23 The data restructure wizard

cases, or cases to variables. We have different levels of time in different columns (variables), and we want them to be in different rows (cases), so we need to select Restructure selected variables into cases. Click on Next > to move to the next dialog box. This dialog box asks you whether you are creating just one new variable in your new data file from different columns in the old data file, or whether you want to create more than one new variable. In our case we are going to create one variable representing life satisfaction; therefore, select One (for example, w1, w2, and w3). When you have done this click on Next >. The next dialog box is crucial because it's where you set up the new data file. By default, SPSS creates a variable in your new data file called **id** which tells you from which person the data came (i.e. which row of the original data file). It does this by using the case numbers in the original data file. This default is fine, but if you want to change it (or the name **id**) then go to the section labelled *Case Group Identification* and change Use case number to be Use selected variable and then select a variable from your data file to act as a label in the new data file. For example, in the diagram I have chosen the variable **Person** from the original data set to identify participants in the new data file.

In the section labelled *Variables to be Transposed* there is a drop-down list labelled *Target Variable* which should contain an item labelled (rather unimaginatively) *trans1*. There is one item because we specified that we wanted one new variable in the previous dialog box (if we had asked for more than one new variable this drop-down list would contain as many items as variables that we requested). We can change the name of *trans1* by selecting the variables in the drop-down list and then editing their names. I suggest that you rename the variable *Life_Satisfaction*. We then need to tell SPSS which columns are associated with these two variables. Select from the list labelled *Variables in the Current File* the four variables that represent the different time points at which life satisfaction was measured (**Satisfaction_Baseline**, **Satisfaction_6_Months**, **Satisfaction_12_Months**, **Satisfaction_18_Months**). If you hold down the *Ctrl* key then you can select all four variables and either drag them across or click on . It's important that you select the variables in the correct order: SPSS assumes that the first variable that it encounters is the first level of the repeated measure, and the second variable is the second level and so on. Once the variables are transferred, you can reorder them by using or to move selected variables up or down the list. Finally, there is a space to select *Fixed Variable(s)*. Drag variables here that do not vary at level 1 of your hierarchy. In this example, this means that we can select variables that are different in different people (they vary at level 2) but have the same value at the different time points (they do not vary at level 1). The only variable that we have like this is **Gender**, which did not change over the course of the data collection, but differs across people. When you have finished click on Next >.

The remaining dialog boxes are fairly straightforward. The next two deal with the indexing variable. SPSS creates a new variable that will tell you from which column the data originate. In our case with four time points this will mean that the variable is simply a sequence of numbers from 1 to 4. So, if the data point came from the baseline phase it will be assigned a 1, whereas it if came from the 18-month follow-up phase it will be given a 4. You can select to have a single index variable (which we need here) or not to have one, or to have several. We have restructured only one repeated-measures variable (**Time**), so we need only one index variable to represent levels of this variable. Therefore, select One and click on Next >. In the next dialog box you can opt either to have the index variable containing numbers (such as 1, 2, 3, 4) or to use the names at the top of the columns from which the data came. The choice is up to you. You can also change the index variables name from *Index* to something useful such as Time (as I have done in the figure). The default options in the remaining dialog boxes are fine for most purposes so you can just click on Finish to do the restructuring. However, if you want to, you can click on Next > to move to a dialog box that enables you to opt to keep any variables from your original data file that you haven't explicitly specified in the previous dialog boxes. The default is to drop them, and that's fine (any variables from the original data file that you want in the new data file should probably be specified earlier on in the wizard). You can also choose to keep or discard missing data. Again, the default option to keep missing data is advisable here because multilevel models can deal with missing values. Click on Next > to move on to the final dialog box. This box gives you the option to restructure the data (the default) or to paste the syntax into a syntax window so that you can save it as a syntax file. If you're likely to do similar data restructuring then saving the syntax might be useful, but once you have got used to the windows it doesn't take long to restructure new data anyway. Click on Finish to restructure the data.

The restructured data are shown in Figure 19.24; it's useful to compare the restructured data with the old data file in Figure 19.22. Notice that each person is now represented by four rows (one for each time point) and that variables such as gender that are invariant over the time points have the same value within each person. However, our outcome variable (life satisfaction) does change over the four time points (the four rows for each person).

There is only one other thing left to do. In your data set you'll notice that the time points have values from 1 to 4. However, it's useful to centre this variable at 0 because our initial life satisfaction was measured before the new relationship. Therefore, an intercept of 0 is

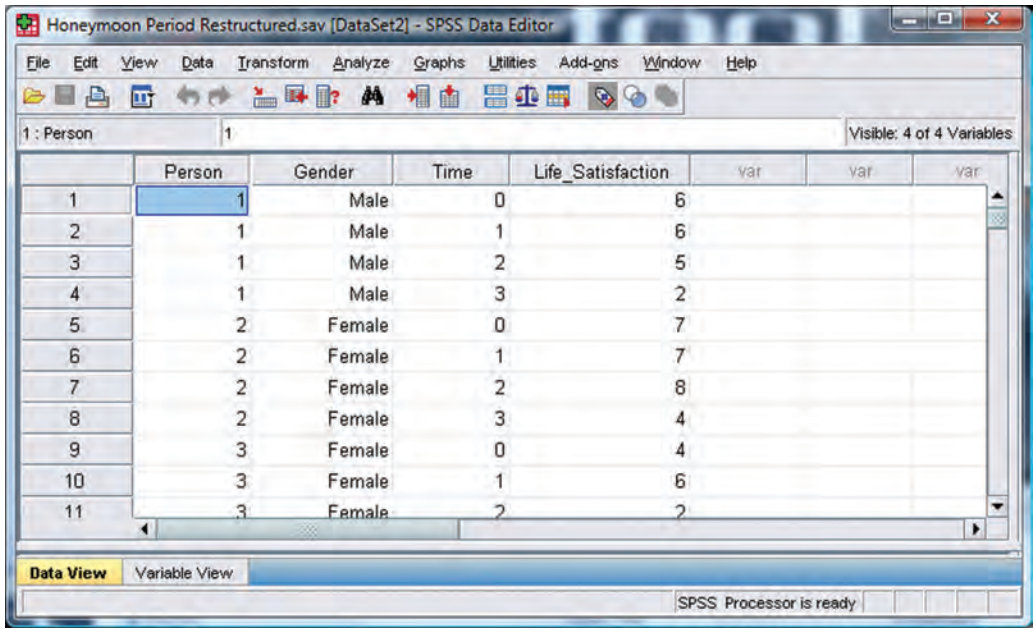


FIGURE 19.24 Data entry for a repeated-measures multilevel model

meaningful for these data: it is the value of life satisfaction when not in a relationship. By centring the scores around a baseline value of 0 we can interpret the intercept much more easily and intuitively. The easiest way to change the values is using the *compute* command to recompute **Time** to be **Time** –1. This will change the values from 1–4 to 0–3.

SELF-TEST Use the *compute* command to transform **Time** into **Time** minus 1.

19.7.4. Running a growth model on SPSS ④

Now that we have our data set up, we can run the analysis. Essentially, we can set up this analysis in a very similar way to the previous example. First, select Analyze Mixed Models Linear... and in the initial dialog box set up the level 2 variable. In this example, life satisfaction at multiple time points is nested within people. Therefore, the level 2 variable is the person and this variable is represented by the variable labelled **Person**. Select this variable and drag it to the box labelled *Subjects* (or click on), see Figure 19.25. Click on Continue to access the main dialog box.

In the main dialog box we need to set up our predictors and outcome. The outcome was life satisfaction, so select **Life_Satisfaction** and drag it to the box labelled *Dependent Variable* (or click on). Our predictor, or growth variable, is **Time** so select this variable and drag it to the box labelled *Covariate(s)*, or click on , see Figure 19.26.

We need to add the potential growth curves that we want to test as fixed effects to our model, so click on Fixed... to bring up the fixed effects dialog box (Figure 19.27). In section 19.7.1 we discussed different growth curves. With four time points we can fit up to a third-order polynomial. One way to do this would be to start with just the linear effect (**Time**), then run a new

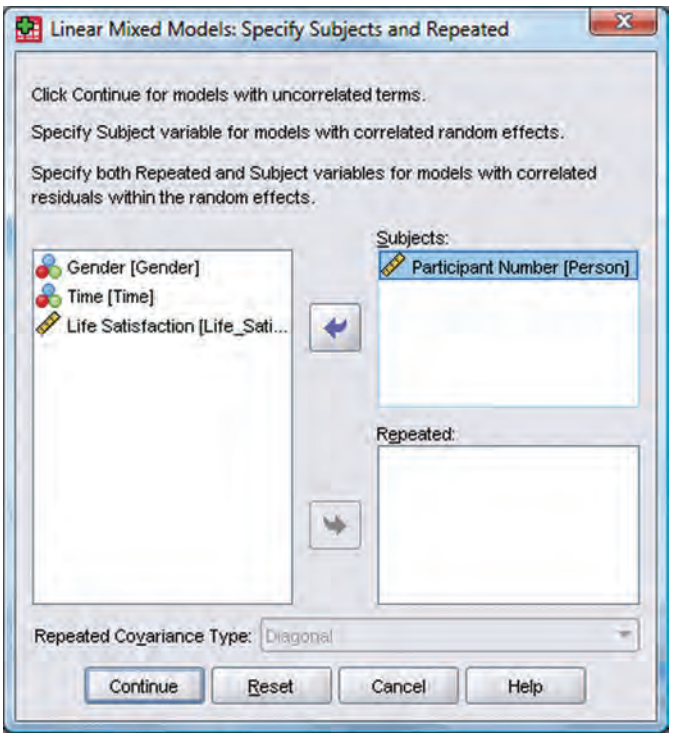


FIGURE 19.25
Setting up the level 2 variable in a growth model

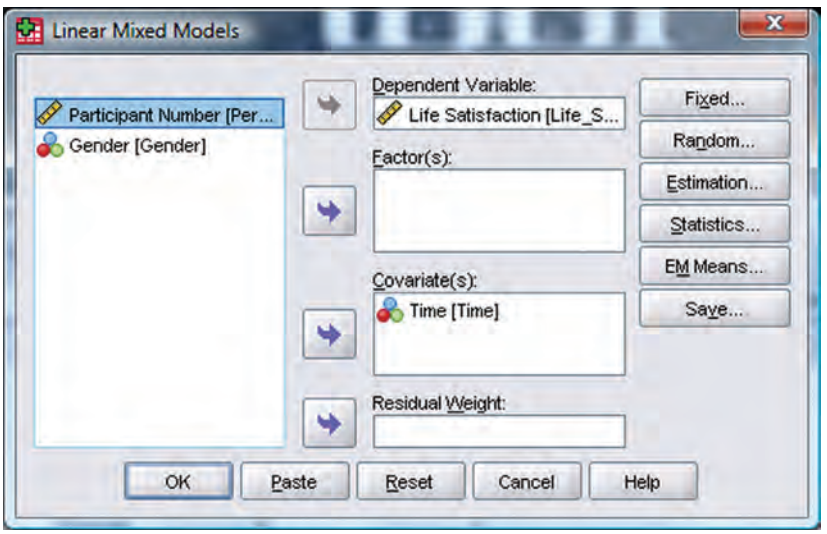


FIGURE 19.26
Setting up the outcome variable and predictor in a multilevel growth model

model with the linear and quadratic (**Time**^{2}) polynomials to see if the quadratic trend improves the model. Finally, run a third model with the linear, quadratic and cubic (**Time**^{3}) polynomial in, and see if the cubic trend adds to the model. So, basically, we add in polynomials one at a time and assess the change in the –2LL. To specify the linear polynomial click on **Time** and then click Add to add it into the model. Click on Continue to return to the main dialog box.

I mentioned earlier on that we expected the relationship between time and life satisfaction to have both a random intercept and a random slope. We need to define these parameters now by clicking on Random... in the main dialog box to access the dialog box in Figure 19.28. The first thing we need to do is to specify our contextual variable. We do this by selecting it from the list

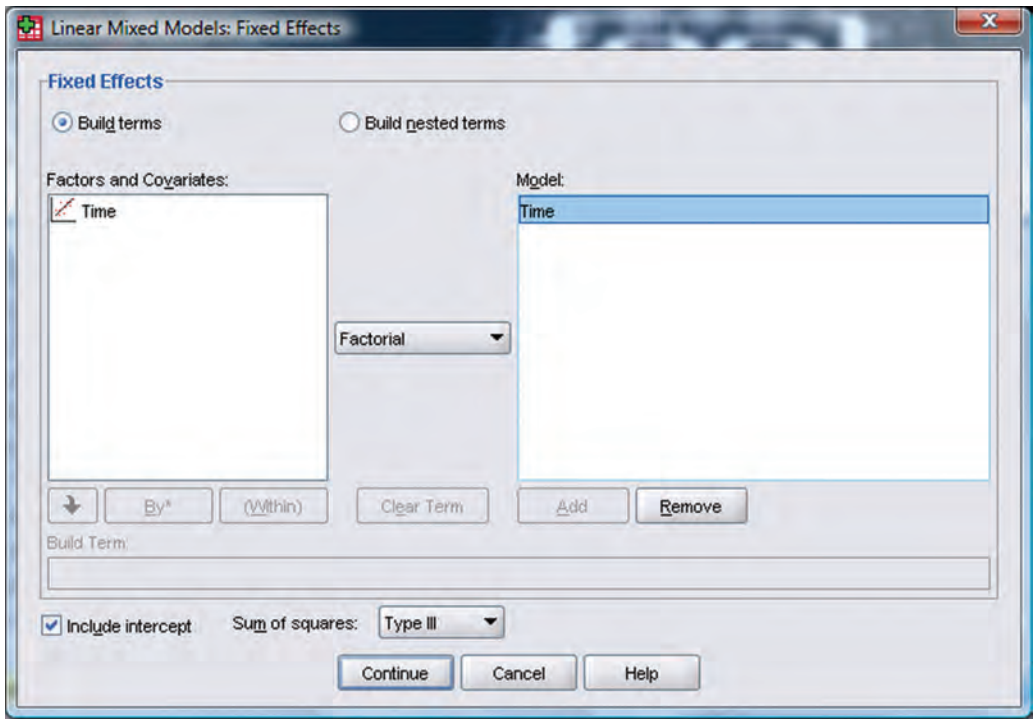


FIGURE 19.27
Setting up the linear polynomial

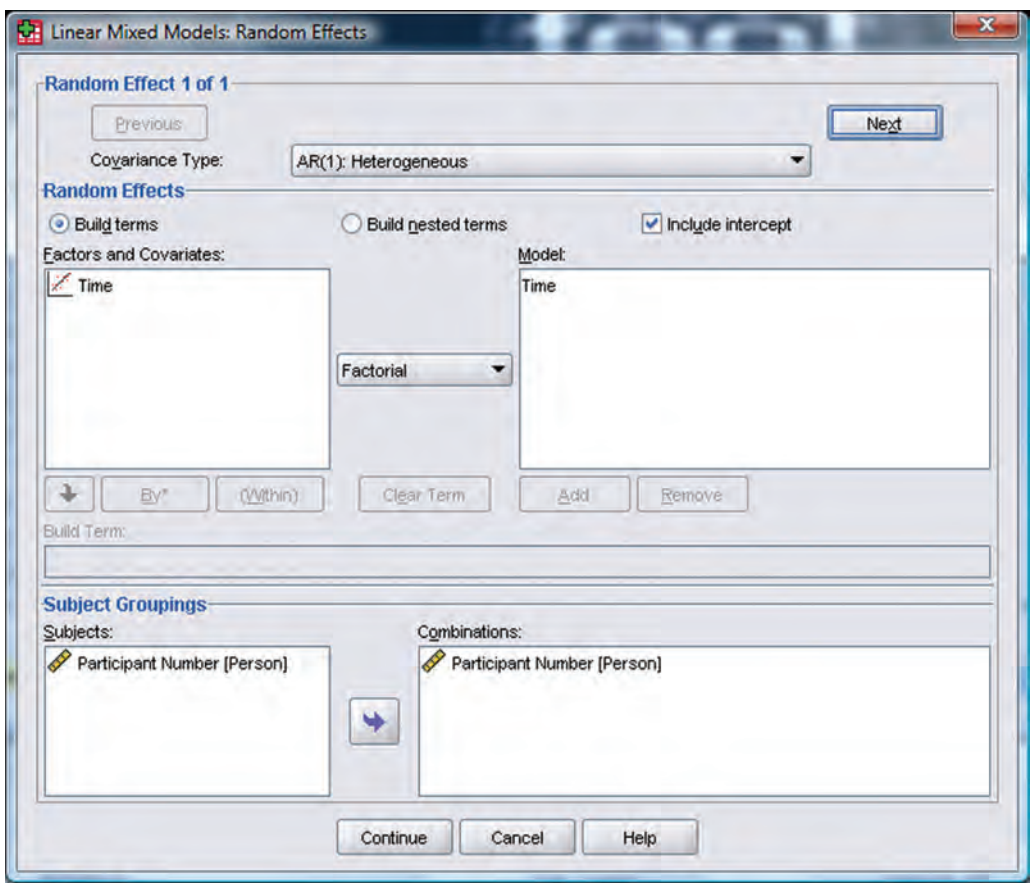


FIGURE 19.28
Defining a random intercept and random slopes in a growth model

of contextual variables that we have told SPSS about already. These appear in the section labelled *Subjects* and because we specified only one variable, there is only one variable in the list, **Person**. Select this variable and drag it to the area labelled *Combinations* (or click on). To specify that the intercept is random select Include intercept, and to specify random slopes for the effect of **Time**, click on this variable in the *Factors and Covariates* list and then click on Add to include it in *Model*. Finally, we need to specify the covariance structure. By default, the covariance structure is set to be Variance Components. However, we saw in section 19.4.2 that when we have repeated measures over time it can be useful to specify a covariance structure that assumes that scores become less correlated over time. Therefore, let's choose an autoregressive covariance structure, AR(1), and let's also assume that variances will be heterogeneous. Therefore, select AR(1): Heterogeneous from the drop-down list. Click on Continue to return to the main dialog box.

Click on Estimation... and select Maximum Likelihood (ML) and then click on Statistics... and select *Parameter estimates* and *Tests for covariance parameters* (see Figure 19.11). Click on Continue to return to the main dialog box. To run the analysis, click on OK.

SPSS Output 19.10 shows the preliminary tables from the output. We can see that the linear trend was significant, $F(1, 106.72) = 134.26$, $p < .001$. For evaluating the improvement in the model when we add in new polynomials, we also need to note the value –2LL, which is 1862.63, and the degrees of freedom, which are 6 (look at the row labelled *Total* in the column labelled *Number of Parameters*, in the table called **Model Dimension**).

Now, let's add the quadratic trend. To do this we return to the dialog box for fixed effects. Therefore, follow the instructions to run this analysis again until you reach the point where you click on Fixed... . The linear polynomial should already be specified from before (if not, then click on **Time** and then click on Add to add it into the model) and the dialog box will look like Figure 19.27. To add the higher-order polynomials we need to select Build nested terms. Select **Time** in the *Factors and Covariates* list and will become active; click on this button and **Time** will appear in the space labelled *Build Term*. For the quadratic or second-order polynomial we need to define **Time**^{2} (**Time** multiplied by itself), and we can specify this by clicking on BY* to add a multiplication symbol to our term, then selecting **Time** again and clicking on . The *Build Term* bar should now read

SPSS OUTPUT 19.10

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	Time	1		1	
Random Effects	Intercept + Time	2	Heterogeneous First-Order Autoregressive	3	Person
Residual				1	
Total		4		6	

a. Dependent Variable: Life Satisfaction.

Information Criteria[a]

-2 Log Likelihood	1862.626
Akaike's Information Criterion (AIC)	1874.626
Hurvich and Tsai's Criterion (AICC)	1874.821
Bozdogan's Criterion (CAIC)	1905.119
Schwarz's Bayesian Criterion (BIC)	1899.119

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Life Satisfaction.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	113.653	1137.088	.000
Time	1	106.715	134.264	.000

a. Dependent Variable: Life Satisfaction.

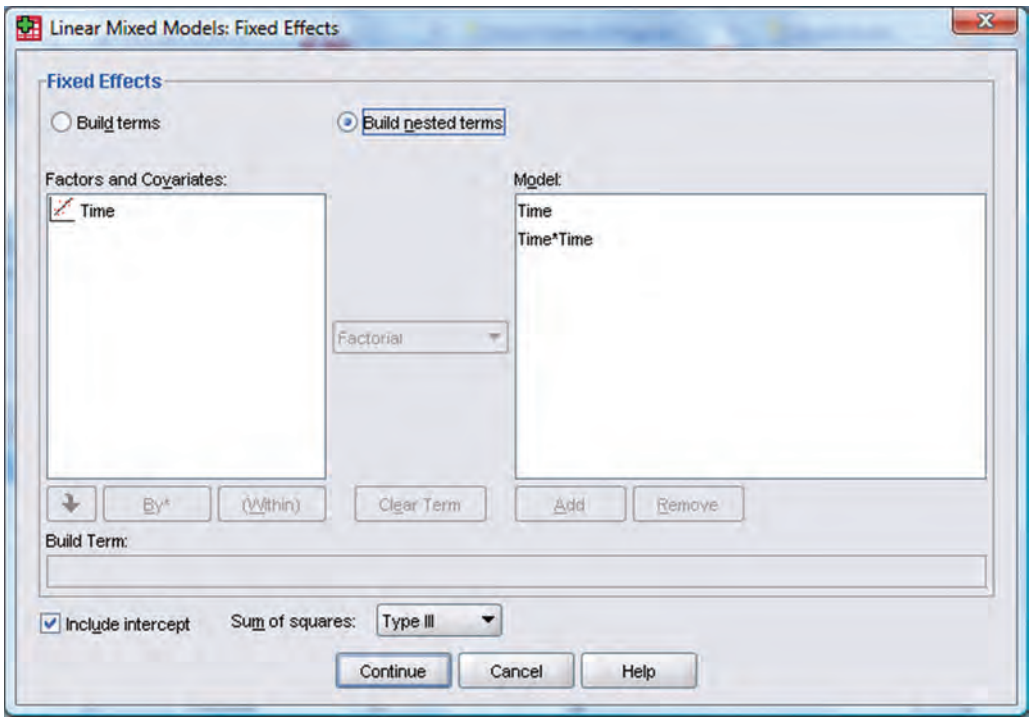


FIGURE 19.29 Specifying a linear trend (*Time*) and a quadratic trend (*Time*Time*)

*Time*Time* (or, put another way, **Time**^{2}). This term is the second-order polynomial, and we click on Add to put it into the model. Click on Continue to return to the main dialog box and click on OK to rerun the analysis.

The output will now include the quadratic polynomial. To see whether this quadratic trend has improved the model we need to compare the –2LL for this new model, to the value when only the linear polynomial was included. The value of –2LL is shown in SPSS Output 19.11, and it is 1802.03. We have added only one term to the model so the new degrees of freedom will have risen by 1, from 6 to 7 (you can check that the new degrees of freedom are 7 in the row labelled *Total* in the column labelled *Number of Parameters*, in the table called **Model Dimension**). We can compute the change in –2LL as a result of the quadratic term by subtracting the –2LL for this model from the –2LL for the model with only the linear trend:

$$\chi^2_{\text{Change}} = 1862.63 - 1802.03 = 60.60$$
$$df_{\text{Change}} = 7 - 6 = 1$$

If we look at the critical values for the chi-square statistic for $df = 1$ in Appendix A4, they are 3.84 ($p < .05$) and 6.63 ($p < .01$); therefore, this change is highly significant because 60.60 is bigger than these values.

SPSS OUTPUT 19.11

Information Criteria[a]

-2 Log Likelihood	1802.026
Akaike's Information Criterion (AIC)	1816.026
Hurvich and Tsai's Criterion (AICC)	1816.287
Bozdogan's Criterion (CAIC)	1851.602
Schwarz's Bayesian Criterion (BIC)	1844.602

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Life Satisfaction.

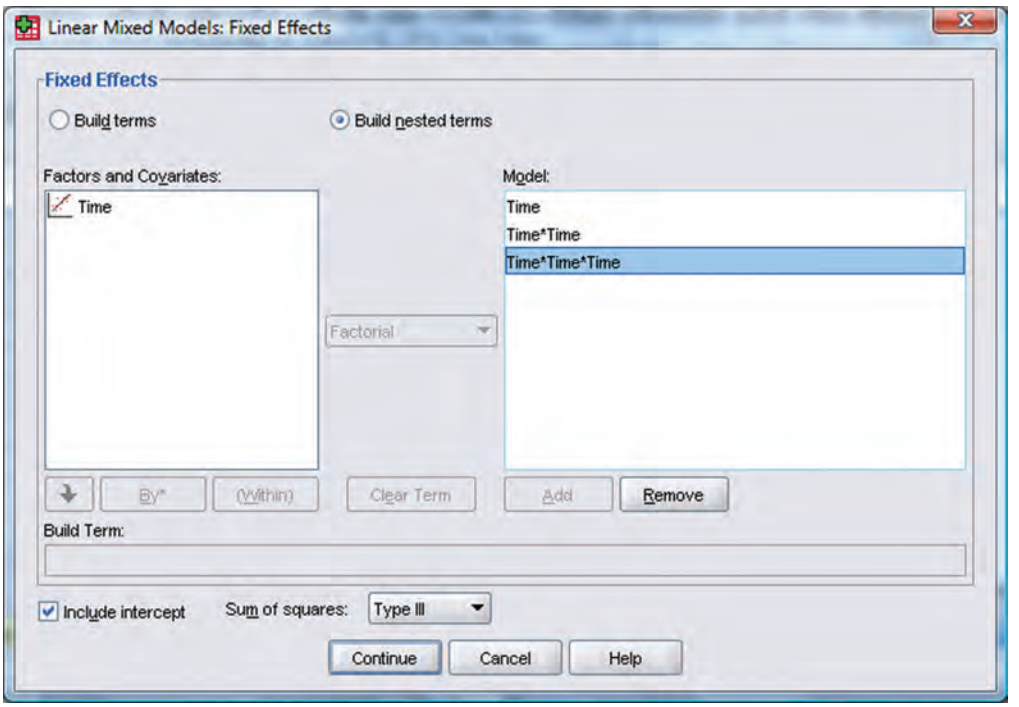


FIGURE 19.30 Specifying linear (*Time*), quadratic (*Time*Time*) and cubic (*Time*Time*Time*) trends

Finally, let's add the cubic trend. To do this we return to the dialog box for fixed effects. As for the quadratic trend, follow the instructions to run this analysis until you reach the point where you click on Fixed... . The linear and quadratic polynomials should already be specified from before and the dialog box will look like Figure 19.29. As for the quadratic trend, make sure Build nested terms is selected. Then, select **Time** in the *Factors and Covariates* list and ➘ will become active; click on this button and **Time** will appear in the space labelled *Build Term*. For the cubic or third-order polynomial we need to define **Time**^{3} (or *Time*Time*Time*). We build this term up in the same way as for the quadratic polynomial. Select **Time**, click on ➘, click on BY* , select **Time** again, click on ➘, click on BY* again, select **Time** for a third time, click on ➘ and finally click on Add . This should add the third-order polynomial (or *Time*Time*Time*) to the model,[8] see Figure 19.30. Click on Continue to return to the main dialog box and click on OK to rerun the analysis.

The output will now include the cubic polynomial. To see whether this cubic trend has improved the model we again compare the –2LL for this new model to the value in the previous model. The value of –2LL is shown in SPSS Output 19.12, and it is 1798.86. We have added only one term to the model so the new degrees of freedom will have risen by 1, from 7 to 8 (again you can find the value of 8 in the row labelled *Total* in the column labelled *Number of Parameters*, in the table called **Model Dimension**). We can compute the change in –2LL as a result of the cubic-term by subtracting the –2LL for this model from the –2LL for the model with only the linear trend:

$$\chi^2_{\text{Change}} = 1802.03 - 1798.86 = 3.17$$
$$df_{\text{Change}} = 8 - 7 = 1$$

[8] Should you ever want even high-order polynomials (notwithstanding my remark about them having little real-world relevance) then you can extrapolate from what I have told you about the other polynomials; for example, for a fourth-order polynomial you go through the whole process again, but this time creating **Time**^{4} (or *Time*Time*Time*Time*), and for the fifth-order polynomial you create **Time**^{5} (or *Time*Time*Time*Time*Time*).

Using the same critical values for the chi-square statistic as before, we can conclude that this change is not significant because 3.17 is less than the critical value of 3.84.

We will look at the SPSS output for this final model in a little more detail (SPSS Output 19.12). First, we are given the fit indices (the –2LL, AIC, AICC, CAIC and BIC). As we have seen, these are useful mainly for comparing models, so we have used the log-likelihood, for example, to test whether the addition of a polynomial significantly affects the fit of the model. The main part of the output is the table of fixed effects and the parameter estimates. These tell us that the linear, $F(1, 221.39) = 10.01$, $p < .01$, and quadratic, $F(1, 212.49) = 9.41$, $p < .01$, trends both significantly described the pattern of the data over time; however, the cubic trend, $F(1, 214.37) = 3.19$, $p > .05$, does not. This confirms what we already know from comparing the fit of successive models. The trend in the data is best described by a second-order polynomial, or a quadratic trend. This reflects the initial increase in life satisfaction 6 months after finding a new partner but a subsequent reduction in life satisfaction at 12 and 18 months after the start of the relationship (Figure 19.21). The parameter estimates tell us much the same thing. It's worth remembering that this quadratic trend is only an *approximation*: if it were completely accurate then we would predict from the model that couples who had been together for 10 years would have negative life satisfaction, which is impossible given the scale we used to measure it.

SPSS OUTPUT 19.12

Information Criteria[a]

-2 Log Likelihood	1798.857
Akaike's Information Criterion (AIC)	1814.857
Hurvich and Tsai's Criterion (AICC)	1815.193
Bozdogan's Criterion (CAIC)	1855.515
Schwarz's Bayesian Criterion (BIC)	1847.515

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: Life Satisfaction.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	137.405	884.469	.000
Time	1	221.392	10.009	.002
Time * Time	1	212.488	9.408	.002
Time * Time * Time	1	214.371	3.187	.076

a. Dependent Variable: Life Satisfaction.

Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	6.634783	.223093	137.405	29.740	.000	6.193644	7.075921
Time	1.544663	.488236	221.392	3.164	.002	.582478	2.506849
Time * Time	-1.323625	.431531	212.488	-3.067	.002	-2.174256	-.472995
Time * Time * Time	.170266	.095374	214.371	1.785	.076	-.017724	.358257

a. Dependent Variable: Life Satisfaction.

Estimates of Covariance Parameters[a]

						95% Confidence Interval	
Parameter		Estimate	Std. Error	Wald Z	Sig.	Lower Bound	Upper Bound
Residual		1.834247	.178865	10.255	.000	1.515144	2.220556
Intercept + Time [subject = Person]	Var: Intercept	3.889341	.700207	5.555	.000	2.732954	5.535028
	Var: Time	.244426	.096858	2.524	.012	.112420	.531437
	ARH1 rho	-.382572	.151351	-2.528	.011	-.635491	-.055507

a. Dependent Variable: Life Satisfaction.

The final part of the output tells us about the random parameters in the model. First of all, the variance of the random intercepts was $\text{Var}(u_{0j}) = 3.89$. This suggests that we were correct to assume that life satisfaction at baseline varied significantly across people. Also, the variance of the people's slopes varied significantly $\text{Var}(u_{1j}) = 0.24$. This suggests also that the change in life satisfaction over time varied significantly across people too. Finally, the covariance between the slopes and intercepts (–0.38) suggests that as intercepts increased, the slope decreased. (Ideally, all of these terms should have been added in individually so that we could calculate the chi-square statistic for the change in the –2LL for each of them.)

19.7.5. Further analysis ④

It's worth pointing out that I've kept this growth curve analysis simple to give you the basic tools. In the example I allowed only the linear term to have a random intercept and slopes, but given that we discovered that a second-order polynomial described the change in responses, we could redo the analysis and allow random intercepts and slopes for the second-order polynomial also. To do these we would just have to specify these terms in Figure 19.28 in much the same way as we set them up as fixed effects in Figure 19.29. If we were to do this it would make sense to add the random components one at a time and test whether they have a significant impact on the model by comparing the log-likelihood values or other fit indices.

Also, the polynomials I have described are not the only ones that can be used. You could test for a logarithmic trend over time, or even an exponential one.

CRAMMING SAM'S TIPS Growth models

- Growth models are multilevel models in which changes in an outcome over time are modelled using potential growth patterns.
- These growth patterns can be linear, quadratic, cubic, logarithmic, exponential, or anything you like really.
- The hierarchy in the data is that time points are nested within people (or other entities). As such, it's a way of analysing repeated-measures data that have a hierarchical structure.
- The **Information Criteria** table can be used to assess the overall fit of the model. The –2LL can be significance tested with *df* = the number of parameters being estimated. It is mainly used, though, to compare models that are the same in all but one parameter by testing the difference in –2LL in the two models against *df* = 1 (if only one parameter has been changed). The AIC, AICC, CAIC and BIC can also be compared across models (but not significance tested).
- The table of **Type III Tests of Fixed Effects** tells you whether the growth functions that you have entered into the model significantly predict the outcome: look in the column labelled *Sig.* If the value is less than .05 then the effect is significant.
- The table labelled **Estimates of Covariance Parameters** tells us about any random effects in the model. These values can tell us how much intercepts and slopes varied over our level 1 variable. The significance of these estimates should be treated cautiously. The exact labelling of these effects depends on which covariance structure you selected for the analysis.
- An autoregressive covariance structure, AR(1), is often assumed in time course data such as that in growth models.

LABCOAT LENI'S REAL RESEARCH 19.1

A fertile gesture ③

Most female mammals experience a phase of 'estrus' during which they are more sexually receptive, proceptive, selective and attractive. As such, the evolutionary benefit to this phase is believed to be to attract mates of superior genetic stock. However, some people have argued that this important phase became uniquely lost or hidden in human females. Testing these evolutionary ideas is exceptionally difficult but Geoffrey Miller and his colleagues came up with an incredibly elegant piece of research that did just that. They reasoned that if the 'hidden-estrus' theory is incorrect then men should find women most attractive during the fertile phase of their menstrual cycle compared to the pre-fertile (menstrual) and post-fertile (luteal) phase.

To measure how attractive men found women in an ecologically valid way, they came up with the ingeneous idea of collecting data from women working at lap-dancing clubs. These women maximize their tips from male visitors by attracting more dances. In effect the men 'try out' several dancers before choosing a dancer for a prolonged dance. For each dance the male pays a 'tip', therefore the more men that choose a particular woman, the more her earnings will be. As such, each dancer's earnings are a good index of how attractive the male customers have found her. Miller and his colleagues argued, therefore, that if women do have an estrus phase then they will be more attractive during this phase and therefore earn more money. This study is a brilliant example of using a real-world phenomenon to address an important scientific question in an ecologically valid way.

The data for this study are in the file **Miller et al. (2007).sav**. The researcher collected data via a website from several dancers (**ID**), who provided data for multiple lap-dancing shifts (so for each person there are several rows of data). They also measured what phase of the menstrual cycle the women were in at a given shift (**Cyclephase**), and whether they were using hormonal contraceptives (**Contraceptive**) because this would affect their cycle. The outcome was their earnings on a given shift in dollars (**Tips**).

A multilevel model can be used here because the data are unbalanced: each woman differed in the number of shifts they provided data for (the range was 9 to 29 shifts); multilevel models can handle this problem.

Labcoat Leni wants you to carry out a multilevel model to see whether **Tips** can be predicted from **Cyclephase**, **Contraceptive** and their interaction. Is the 'estrus-hidden' hypothesis supported? Answers are in the additional material on the companion website (or look at page 378 in the original article).

MILLER, G. TYBUR, J.M. & JORDAN, B.D. (2007). *EVOLUTION AND HUMAN BEHAVIOR, 28*, 375–381.

19.8. How to report a multilevel model ③

Specific advice on reporting multilevel models is hard to come by. Also, the models themselves can take on so many forms that giving standard advice is hard. If you have built up your model from one with only fixed parameters to one with a random intercept, and then random slope, it is advisable to report all stages of this process (or at the very least report the fixed-effects-only model and the final model). For any model you need to say something about the random effects. For the final model of the cosmetic surgery example you could write something like:

✓ The relationship between surgery and quality of life showed significant variance in intercepts across participants, $\text{var}(u_{0j}) = 30.06$, $\chi^2(1) = 15.05$, $p < .01$. In addition, the slopes varied across participants, $\text{var}(u_{1j}) = 29.35$, $\chi^2(1) = 21.49$, $p < .01$, and the slopes and intercepts negatively and significantly covaried, $\text{cov}(u_{0j}, u_{1j}) = -28.08$, $\chi^2(1) = 17.38$, $p < .01$.

For the model itself, you have two choices. The first is to report the results rather like an ANOVA, with the *F*s and degrees of freedom for the fixed effects, and then report the parameters for the

random effects in the text as well. The second is to produce a table of parameters as you would for regression. For example, we might report our cosmetic surgery example as follows:

- ✓ Quality of life before surgery significantly predicted quality of life after surgery, $F(1, 268.92) = 33.65$, $p < .001$, surgery did not significantly predict quality of life, $F(1, 15.86) = 2.17$, $p = .161$, but the reason for surgery, $F(1, 259.89) = 9.67$, $p < .01$, and the interaction of the reason for surgery and surgery, $F(1, 217.09) = 6.28$, $p < .05$, both did significantly predict quality of life. This interaction was broken down by conducting separate multilevel models on the 'physical reason' and 'attractiveness reason'. The models specified were the same as the main model but excluded the main effect and interaction term involving the reason for surgery. These analyses showed that for those operated on only to change their appearance, surgery almost significantly predicted quality of life after surgery, $b = -4.31$, $t(7.72) = -1.92$, $p = .09$: quality of life was lower after surgery compared to the control group. However, for those that had surgery to solve a physical problem, surgery did not significantly predict quality of life, $b = 1.20$, $t(7.61) = 0.58$, $p = .58$. The interaction effect, therefore, reflects the difference in slopes for surgery as a predictor of quality of life in those that had surgery for physical problems (slight positive slope) and those that had surgery purely for vanity (a negative slope).

Alternatively we could present parameter information in a table:

	b	*SE* b	*95% CI*
Baseline QoL	0.31	0.05	0.20, 0.41
Surgery	–3.19	2.17	–7.78, 1.41
Reason	–3.51	1.13	–5.74, –1.29
Surgery × Reason	4.22	1.68	0.90, 7.54

What have I discovered about statistics? ②

Writing this chapter was quite a steep learning curve for me. I've been meaning to learn about multilevel modelling for ages, and now I finally feel like I know something. This is pretty amazing considering that the bulk of the reading and writing was done between 11 p.m. and 3 a.m. over many nights. However, despite now feeling as though I understand them, I don't, and if you feel like you now understand them then you're wrong. This sounds harsh, but sadly multilevel modelling is very complicated and we have scratched only the surface of what there is to know. Multilevel models often fail to converge with no apology or explanation, and trying to fathom out what's happening can feel like hammering nails into your head.

Needless to say I didn't mention any of this at the start of the chapter because I wanted you to read it. Instead, I lulled you into a false sense of security by looking gently at how data can be hierarchical and how this hierarchical structure can be important. Most of the tests in this book simply ignore the hierarchy. We also saw that hierarchical models are just basically a fancy regression in which you can estimate the variability in the slopes and intercepts within entities. We saw that you should start with a model that ignores the hierarchy and then add in random intercepts and slopes to see if they improve the fit of

the model. Having submerged ourselves in the warm bath of standard multilevel models we moved on to the icy lake of growth curves. We saw that there are ways to model trends in the data over time (and that these trends can also have variable intercepts and slopes). We also discovered that these trends have long confusing names like fourth-order polynomial. We asked ourselves why they couldn't have a sensible name, like Kate. In fact, we decided to ourselves that we'd secretly call a linear trend Kate, a quadratic trend Benjamin, a cubic trend Zoë and a fourth-order trend Doug. 'That will show the statisticians' we thought to ourselves, and felt a little bit self-satisfied too.

We also saw that after years of denial, my love of making a racket got the better of me. This brings my life story up to date. Admittedly I left out some of the more colourful bits, but only because I couldn't find an extremely tenuous way to link them to statistics. We saw that over my life I managed to completely fail to achieve any of my childhood dreams. It's OK, I have other ambitions now (a bit smaller scale than 'rock star') and I'm looking forward to failing to achieve them too. The question that remains is whether there is life after *Discovering Statistics*. What effect does writing a statistics book have on your life?

Key terms that I've discovered

AIC
AICC
AR(1)
BIC
CAIC
Centring
Diagonal
Fixed coefficient
Fixed effect
Fixed intercept
Fixed slope
Fixed variable
Grand mean centring
Group mean centring
Growth curve
Multilevel linear model
Polynomial
Random coefficient
Random effect
Random intercept
Random slope
Random variable
Unstructured
Variance components

Smart Alex's tasks

- **Task 1**: Using the cosmetic surgery example, run the analysis described in section 19.6.5 but also including BDI, age and gender as fixed effect predictors. What differences does including these predictors make? ④
- **Task 2**: Using our growth model example in this chapter, analyse the data but include **Gender** as an additional covariate. Does this change your conclusions? ④
- **Task 3: Getting kids to exercise (Hill, Abraham, & Wright, 2007)**: The purpose of this research was to examine whether providing children with a leaflet based on the 'theory of planned behaviour' increases children's exercise. There were four different interventions (**Intervention**): a control group, a leaflet, a leaflet and quiz, and a leaflet and plan. A total of 503 children from 22 different classrooms were sampled (**Classroom**). It was not practical to have children in the same classrooms in different conditions, therefore the 22 classrooms were randomly assigned to the four

different conditions. Children were asked 'On average over the last three weeks, I have exercised energetically for at least 30 minutes _____ times per week' after the intervention (**Post_Exercise**). Run a multilevel model analysis on these data (**Hill et al. (2007).sav**) to see whether the intervention affected the children's exercise levels (the hierarchy in the data is: children within classrooms within interventions). ④

- **Task 4**: Repeat the above analysis but include the pre-intervention exercise scores (**Pre_Exercise**) as a covariate. What difference does this make to the results? ④

Answers can be found on the companion website.

Further reading

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage. (This is a fantastic book that is easy to get into but has a lot of depth too.)

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon. (Chapter 15 is a fantastic account of multilevel linear models that goes a bit more in depth than I do.)

Twisk, J. W. R. (2006). *Applied multilevel analysis: a practical guide*. Cambridge: Cambridge University Press. (An absolutely superb introduction to multilevel modelling. This book is exceptionally clearly written and is aimed at novices. Without question, this is the best beginner's guide that I have read.)

Online tutorial

The companion website contains the following Flash movie tutorial to accompany this chapter:

- Mixed Models using SPSS

Interesting real research

Cook, S. A., Rosser, R., & Salmon, P. (2006). Is cosmetic surgery an effective psychotherapeutic intervention? A systematic review of the evidence. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, *59*, 1133–1151.

Miller, G., Tybur, J. M., & Jordan, B. D. (2007). Ovulatory cycle effects on tip earnings by lap dancers: economic evidence for human estrus? *Evolution and Human Behavior*, *28*, 375–381.